
ЭКСПЕРТИЗА ПЕРЬЕВОГО МАТЕРИАЛА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ

О.Л. Силаева¹, А.Н. Варакин², В.Д. Ильичёв³

^{1,3}Учреждение Российской Академии наук
Институт проблем экологии и эволюции им. А.Н. Северцова РАН
Ленинский пр., 33, Москва, Россия, 119071

²Научно-исследовательский центр распознавания образов
ул. Удальцова, 85, Москва, Россия, 119607

Приведены результаты экспертизы перьевого материала с помощью анализа многомерных данных. Цель — создание электронной системы таксономической идентификации птиц по перьевому материалу.

Ключевые слова: таксономическое определение птиц по перу, птерилография, анализ многомерных данных.

В исследованиях пера заинтересована не только фундаментальная наука. Птерилография применяется в практических и теоретических сферах деятельности человека, связанных с орнитологическими дисциплинами, находящихся на стыке с орнитологией, а также удаленных от нее. В первую очередь в практических птерилографических исследованиях заинтересована авиационная орнитология. Определение таксономической и популяционной принадлежности птицы необходимо для анализа столкновений летательных аппаратов с птицами и рекомендаций по их предотвращению. Определение птицы по перу найдет применение в исследованиях разных сторон биологии животных: биоразнообразия; миграций и линьки птиц; внутривидовой и фенотипической изменчивости и разнообразных адаптаций.

Идентификация птиц по перу проводится также при решении задач палеонтологии и палеобиологии, археологии, этнографии и антропологии, в криминалистических исследованиях и при проведении биологической экспертизы. Методика определения видовой принадлежности птицы по перу интересует орнитологов и любителей-натуралистов. Все это станет возможным при наличии электронного определителя (классификатора) нового поколения с автоматической идентификацией таксона птицы по первостепенному маховому перу (1).

В последние десятилетия созданы мультимедийные определители по перьевым останкам птиц [2—4]. Однако работа с такими определителями требует знаний и навыков как в области полевой орнитологии, так и в работе с электронными носителями и идентификационным ключом.

Материал и методика исследования. Сотрудники Лаборатории экологии и управления поведением птиц ИПЭЭ РАН совместно со специалистами Научно-исследовательского центра распознавания образов впервые применили методы анализа многомерных данных для изучения перьевого материала [1]. В частно-

сти, использовали кластерный анализ и метод главных компонент с целью выявить связи и группы схожих объектов, имеющиеся в массиве перьевых данных, а также обнаружить скрытые закономерности обучающей выборки.

Объект, т.е. полное перо птицы, описали набором значимых характеристик, цифровых признаков. Биометрические измерения проводились в программе Corel Draw. В настоящее время используется оригинальная программа, разработанная В.А. Никулиным PteroMetr 1 (β-версия).

Из эталонной базы данных по макроструктуре пера были выбраны четыре дигитальные ПМ, т.е. со второго по пятое перо. Перья при необходимости промыли в Triton-X-100, сушили и сканировали с разрешением 600 dpi. По каждому перу первоначально было сделано три измерения (рис. 1). Перья, принадлежавшие 18 особям 7 видам птиц, положили начало малой обучающей выборке (выборка 1) для синтеза идентификационного набора признаков. Таким образом, каждая особь характеризовалась 12 признаками (измерениями).

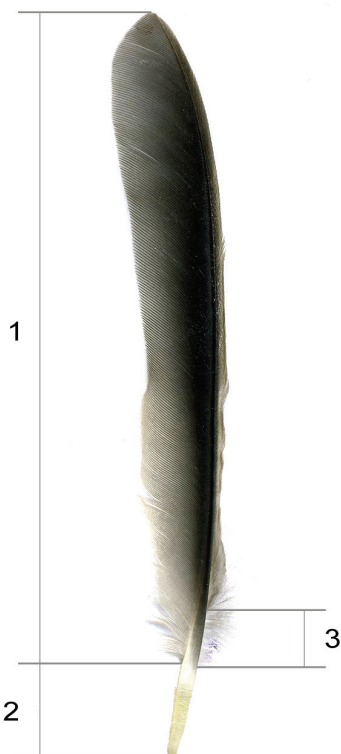


Рис. 1. Промеры пера для выборки 1 (мм)

Результаты и обсуждение. С помощью программы кластерного анализа было проведено исследование с целью выявления пространства, в котором каждый таксон группировался бы таким образом, чтобы можно было набором прямых отделить имеющиеся кластеры друг от друга. При проектировании из 12-мерного пространства на двухмерную плоскость необходимо было подобрать такие проекции, при которых кластер, описывающий особь в виде признаков, группируется и отде-

ляется от всех остальных кластеров. При различных проекциях кластеры могут пересекаться. Однако если хотя бы в одной проекции кластер отделяется от других, то необходимое условие для таксономической идентификации выполняется. Таким образом, в 12-мерном пространстве было найдено такое проективное преобразование на двумерную плоскость, при котором внутрикласовое расстояние одного вида было минимальным, а межклассовое расстояние — максимальным. Для такого типа данных был найден набор прямых, которые вполне четко отделили один вид от другого.

В дальнейшем была собрана более представительная обучающая выборка (выборка 2), она состояла из 31 вида, 14 семейств и 80 особей. По каждому перу было сделано уже 12 измерений, в результате каждая особь характеризовалась 48 признаками (рис. 2).

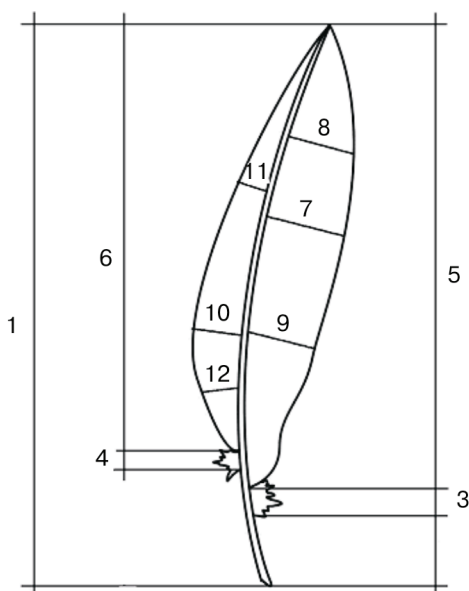


Рис. 2. Промеры пера для выборки 2:

1 — длина ствола; 2 — длина очина; 3 — длина пуховой части внутреннего опахала (2); 4 — длина пуховой части наружного опахала; 5 — длина контурной части внутреннего опахала; 6 — длина контурной части наружного опахала; 7 — максимальная ширина внутреннего опахала; 8 — дистальная ширина внутреннего опахала; 9 — проксимальная ширина внутреннего опахала; 10 — максимальная ширина наружного опахала; 11 — дистальная ширина наружного опахала; 12 — проксимальная ширина наружного опахала

Номера измерений соответствуют номерам признаков, например: 1 — длина ствола — P1 и т.д. (см. рис. 2; табл. 1, 2).

Определение информативности и корреляции признаков. Априори известно, что коррелированные признаки малоинформативны. По всем четырем исследованным перьям и 48 признакам был проведен анализ пар информативности признаков и их корреляции. Большинство признаков показало высокую информативность. Исключением стал третий признак, а именно длина пуховой части внутреннего опахала (табл. 1), показавший наименьшую информативность у всех четырех ПМ. При этом именно этот рисунок обладал низкой корреляцией (табл. 2).

Таблица 1

Информативность признаков выборки 2

Перо	Признак											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
ПМ2	0,7694	0,7570	0,1457	0,5938	0,7681	0,7748	0,7392	0,7882	0,7558	0,6420	0,6214	0,6390
ПМ3	0,8118	0,7824	0,3988	0,7161	0,8132	0,8047	0,7463	0,7974	0,7132	0,6632	0,5282	0,4186
ПМ4	0,8294	0,7946	0,4345	0,6585	0,8321	0,8250	0,7387	0,8290	0,7168	0,7306	0,4560	0,6984
ПМ5	0,8404	0,8069	0,4059	0,6921	0,8421	0,8443	0,7677	0,8176	0,7325	0,7260	0,4315	0,6507

Таблица 2

Корреляция признаков для ПМ2*

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	1,000											
P2	0,987	1,000										
P3	-0,192	-0,238	1,000									
P4	0,871	0,875	-0,241	1,000								
P5	0,999	0,983	-0,195	0,869	1,000							
P6	0,997	0,979	-0,184	0,842	0,998	1,000						
P7	0,981	0,981	-0,210	0,873	0,979	0,974	1,000					
P8	0,958	0,950	-0,209	0,842	0,957	0,954	0,961	1,000				
P9	0,969	0,966	-0,217	0,840	0,968	0,965	0,979	0,956	1,000			
P10	0,906	0,917	-0,234	0,839	0,902	0,894	0,914	0,866	0,878	1,000		
P11	0,815	0,837	-0,263	0,710	0,812	0,808	0,829	0,827	0,839	0,870	1,000	
P12	0,899	0,914	-0,246	0,833	0,895	0,888	0,906	0,841	0,866	0,947	0,833	1,000

*Данные симметричны относительно главной диагонали, поэтому заполнена лишь половина таблицы.

Информативность признака P3 для ПМ2 минимальна. Коэффициент корреляции между признаком P1 и P3 наименьший — 0,192 (см. табл. 1). Скорее всего, низкая информативность данного признака связана с тем, что длина пуховой части внутреннего опахала нестабильна и колеблется в широких пределах.

Дальнейшее расширение обучающей выборки и ее анализ показали, что в зависимости от целевой установки данные обучающей выборки могут группироваться по-разному. С этой целью было введено понятие сегментации.

Методика сегментации. В состав обучающей выборки (выборка 3) входило два отряда, 14 семейств, 25 родов, 34 видов и 119 особей сегментации производилась по отрядам, семействам и родам. Предварительно были проведены интегральные оценки вариативности всех ПМ по 12 признакам. Интегральная оценка исследованных перьев показала, что они очень мало отличаются друг от друга. Следовательно, идентификация может проходить по одному ПМ. ПМ2 во всех исследованных группах птиц оказалось наиболее вариативно по выбранным признакам.

Сегментация по отрядам, т.е. по наиболее высоким таксонам получилась четкой — воробьинообразные отделились от голубеобразных. Правда, следует отметить, что воробьинообразные были представлены значительным количеством таксонов, а голубеобразные — только одним таксоном, одним видом сизого голубя. Информативность 3-го (P3) и 11-го (P11) признаков позволила разделить отряды воробьинообразных и голубеобразных. Было показано, что эти два признака достаточно характеризуют данные отряды по каждому из перьев в отдельности. Отряды могут быть разделены и по другим признакам, однако наибольшее межклассовое расстояние показала вышеназванная пара признаков. При сегментации по семейст-

вам в отряде воробьинообразных четко выявились врановые и дроздовые. Сегментация по родам выявила следующие таксоны: *Bombucilla*, *Turdus* (внутри рода рябинника), *Corvus* (распался на две группы) и *Emberiza*.

На основании полученных результатов можно предположить, что на эффективность работы алгоритмов классификации существенным образом будет влиять размер пера. Фактически размер пера — это мультипликативная константа, с которой связаны многие линейные параметры пера. Минимизация влияния мультипликативных констант при создании классификатора в настоящий момент является актуальной задачей, которую предстоит решить.

Центроиды. Дальнейшее увеличение базы данных привело к тому, что визуализация кластеров на двухмерной плоскости стала проблематичной, поэтому от кластеров мы перешли к центроидам, которые характеризуют центр тяжести кластера. Такой переход позволил визуализировать и исследовать имеющуюся в настоящий момент обучающую выборку, однако очевидно, что ее дальнейшее увеличение потребует разработки формальных методов оценки информативности признаков на сколь угодно значительной обучающей выборке.

На основании ранее проведенной сегментации было проведено исследование распределения центроидов внутри семейств и внутри родов с целью определения пар информативных признаков, которые бы удовлетворительно разделяли соответствующие таксоны. Благодаря использованию программы в среде MATLAB были выявлены следующие пары признаков: 1\9; 2\9; 4\9; 6\8; 5\8; 4\8; 2\8; 5\7; 4\7; 2\7; 4\6; 7\6; 4\5. В пространстве пар признаков было получено распределение всех особей по родам в двухмерном пространстве.

Следует заметить, что третий признак (P3) не встречается в парах и не участвует в разделении таксонов. Здесь мы снова столкнулись с малой информативностью этого признака. Однако нестабильностью отличается и четвертый признак (P4). При этом он пять раз встречается в парах.

Таким образом, в имеющейся обучающей выборке были найдены информативные пары признаков, которые показывают принципиальную возможность разделения материала на интересующие нас классы объектов в зависимости от сегментации. Вместе с тем для определения видовой принадлежности необходима более представительная обучающая выборка, содержащая не менее 30 особей каждого вида. Для ее обработки требуются формальные методы определения информативности признаков, которые должны быть инвариантны к мультипликативным константам. При этом с уверенностью можно сказать уже сейчас, что полученные нами результаты, свидетельствуют о том, что достаточно одного из четырех исследованных ПМ для идентификации вида птицы.

Авторы выражают благодарность В.А. Никулину, Н.Ю. Сапунковой и П.Г. Полежанкиной за инженерную и техническую помощь, а также всем, кто помогал в сборе материала.

ПРИМЕЧАНИЯ

- (1) Работы ведутся и по созданию аналогичного микроструктурного определителя. Определение таксона будет возможно при наличии частицы пера с пуховой структурой.
- (2) При отсутствии пуховой части опахала ее длина принималась равной 0,001 мм.

ЛИТЕРАТУРА

- [1] Силаева О.Л., Вараксин А.Н., Ильичёв В.Д. Таксономическая идентификация по перу с помощью кластерного анализа // Актуальные проблемы экологии и природопользования. Вып. 12: Сборник научных трудов. — М.: Луч, 2010. С. 191—196.
- [2] Силаева О.Л., Ильичёв В.Д., Чернова О.Ф., Фадеева Е.О. Мультимедийный определитель птиц по перу и его фрагментам. Отряд воробьинообразные (Passeriformes). Семейство врановые (Corvidae). Компакт-диск // ИПЭЭ им. А.Н. Северцова РАН. 2010.
- [3] Prast W., Shamoun J., Bierhuizen B., Roselaar C.S., Schalk P.H., Wattel J., Los W., Leshem Y., Yom-Tov Y., Buurma L.S. A computer based bird remains identification system (BRIS). Birds of Europe. CD-ROM, Amsterdam: ETI, 1996.
- [4] Prast W., Shamoun J. Bird Remains Identification System (BRIS). Birds of Europe. CD-ROM, Amsterdam: ETI, 2001.

EXAMINATION OF FEATHER MATERIAL WITH THE USE OF MULTIDIMENSIONAL DATA ANALYSIS

O.L. Silaeva¹, A.N. Varaksin², V.D. Ilyichev³

^{1,3}Institution of Russian Academy of Sciences Severtsov
Institute of Ecology and Evolution RAS
Leninsky pr., 33, Moscow, Russia, 119071

²Pattern recognition research company
Udalzov str., 85, Moscow, Russia, 119607

The results of feather expertise by means of multidimensional data analysis are given. The aim is to create the electronic system of taxonomic birds' identification by feather.

Key words: birds' identification by feather, pterylography, multidimensional data analysis.