

КРАТКИЕ НАУЧНЫЕ СООБЩЕНИЯ

УДК 578.087.1

ПАВЕЛ П. ГАМБАРЯН

ВЗВЕШИВАНИЕ ПРИЗНАКОВ И ИХ КОРРЕЛЯЦИЯ

Числовая таксономия решает в основном 2 задачи: а) создание диагностических систем для различения таксонов по комбинациям признаков; б) классификация, т. е. разделение множества таксонов на иерархию более сходных групп. Если число таксонов n , то число попарных сравнений их равно $\frac{n(n-1)}{2}$. Число отличающихся друг от друга объектов при помощи одного ряда взаимоисключающих признаков [10] равно $\sum_{i=1}^m n_i n_j$, $i=1\dots m$, $j=2\dots m$, $i \neq j$, где m —число признаков в ряду, n_i —число таксонов с i -тым признаком. Значение ряда признаков для диагностики можно определить числом отличающихся друг от друга таксонов. Для классификации ценность ряда признаков можно определить числом измененных оценок сходства, которое дает данный ряд признаков (табл. 1, 2).

Таблица 1

6 рядов признаков и число отличающихся друг от друга объектов

№ признака	№ признаков					
	1	2	3	4	5	6
1	б	а	а	а	а	а
2	б	б	а	б	б	б
3	б	б	б	б	в	в
4	б	б	б	в	в	г
5	б	б	б	в	в	д
Число отличий	0	4	6	8	7	10

Таблица 2

Различия объектов X_0 и измененные оценки различия при отбрасывании поочередно по одному признаку X_1-6 из табл. 1

№ объектов X_0	№ признаков					
	X_1	X_2	X_3	X_4	X_5	X_6
1	2	4	4	3	4	3
1	3	5	5	4	4	4
1	4	5	5	4	4	4
1	5	5	5	4	4	4
2	3	3	3	2	3	2
2	4	4	4	3	3	3
2	5	4	4	4	3	3
3	4	2	2	2	1	2
3	5	2	2	2	1	2
4	5	1	1	1	1	1
Число изменений	0	4	6	8	7	10

Число измененных оценок сходства (различий) также равно $\sum n_i n_j$. При сравнении пары таксонов мы используем не всю информацию признаков ряда, а только $\frac{2}{m}$ ее часть. Вес сравнения таксонов по данному

ряду признаков нами оценивается $\frac{2\sum n_i n_j}{m}$.

До сих пор [6, 7, 12, 14, 15], говоря о взвешивании признаков, мы подразумевали их приведенными к биномиальному виду—теза «признак а»

с частотой $\frac{n_1}{n} = p$ и антитеза «не признак а» с частотой $\frac{n_2}{n} = q$, $p + q = 1$.

Для биномиальных признаков нами было предложено оценивать признак произведением частот тезы и антитезы pq [6, 7]. Эта оценка биномиального признака является частным случаем более общей оценки $\frac{2 \sum n_i n_j}{m}$.

Если $m=2$, то вес сравнения объектов равен $n_1 n_2$, $pq \cdot n^2 = n_1 n_2$. Некоторые исследователи [12—16] предпочитали взвешивать признаки по методу Смирнова [12], другие [11, 17] — вообще не взвешивать признаки. Но ряды признаков явно неравнозначны. Нам удалось показать [9], что как метод Смирнова, так и метод разнотипной оценки всех признаков уступают в образовании межгрупповой дисперсии методу с оценкой признака pq . Вывод Тамарина [14—16] о том, что метод Смирнова единственно приемлем, ошибочен. Максимальная оценка редкого признака неверна потому, что этот признак мало влияет на матрицу оценок сходства в силу своей редкости.

Если признаки образуют ранжированный ряд с начальным нулевым рангом (баллом), когда первый ранг ближе ко второму, чем к третьему, или количественный признак выражен через ранжированный ряд с начальным нулевым рангом, оценка $\frac{2 \sum n_i n_j}{m}$ неприменима. В таком случае сумму рангов всех таксонов $\sum R$ делим на произведение максимального ранга и числа объектов $R_{\max} \cdot n$, определяя ценность ранжированного признака величиной $\frac{\sum R}{R_{\max} \cdot n} \left(1 - \frac{\sum R}{R_{\max} \cdot n}\right)$, аналогичной величине pq . Сравнение двух объектов из ранжированного ряда оцениваем $\frac{R_1 - R_2}{R_{\max}} \cdot pq$ [1, 8]. Чтобы эту величину сделать сравнимой с численностями, умножаем ее на n^2 .

Сравнение таксономического и диагностического значения отдельных рядов признаков, веса признаков исчерпывается приведенными оценками. Но числовая таксономия не ограничивается использованием одного ряда признаков. Суммировать веса совпадающих признаков для определения сходства (или несовпадающих признаков для определения различия) таксонов можно только в случае независимости признаков, когда один ряд признаков совершенно не дублирует информации другого ряда признаков. Но так как такого ряда признаков нет, надо учитывать коррелированность, сопряженность признаков. Разные коэффициенты сопряженности, коэффициенты корреляции для измерения зависимости двух альтернативных признаков [2—5, 13, 17] в этом случае мало подходят, так как не рассчитаны на сочетания ранжированных и неранжированных рядов с разным числом признаков. Многомерные статистики с обращением корреляционных матриц тоже не применимы, так как не разработано приемлемого метода определения корреляции рядов признаков.

Для диагностики необходимо и достаточно, чтобы n таксонов отли-

чались друг от друга во всех $\frac{n(n-1)}{2}$ сравнениях. Каждую неповторяющуюся комбинацию признаков будем считать за признак ряда. Тогда число отличий будет $\sum n_i n_j$. Если $\sum n_i n_j = \frac{n(n-1)}{2}$, то все объекты можно отличить друг от друга, и добавление новых признаков будет дублированием информации, так как новой информации оно не дает.

Для классификации, кроме установления различия таксонов, необходимо учитывать и величину этого различия. Воспользуемся тем, что уменьшение всех оценок сходства (или различия) на одну и ту же величину не влияет на результаты классификации. Поэтому для установления сходства после подсчета суммы весов всех совпадений при попарном сравнении всех таксонов (или суммы весов несовпадающих признаков для установления различия) надо из всех оценок сходства (различия) вычесть минимальное значение. Этим будет устранено влияние коррелированности признаков, дублирование их информации.

Обобщенная оценка веса сравнения таксонов $\frac{2 \sum n_i n_j}{m}$ позволяет не приводить ряды взаимоисключающих признаков к биномиальному виду, что ведет к компактности таблиц распределения признаков.

Для целей диагностики сведение неповторяющихся комбинаций признаков к одному ряду позволяет учесть коррелированность признаков, не вычисляя ее.

Для целей классификации вычитание из всех оценок сходства (различия) их минимального значения позволяет учесть коррелированность, дублирование информации и в этом случае, не вычисляя коэффициентов корреляции.

Институт ботаники
АН АрмССР

Поступило 21.III 1973 г.

ՊԱՎԵԼ Պ. ՂԱՄԲԱՐՅԱՆ

ՀԱՏԿԱՆԻՇՆԵՐԻ ՔԱՇԸ ԵՎ ԿՈՌԵԼԱՑԻԱՆ

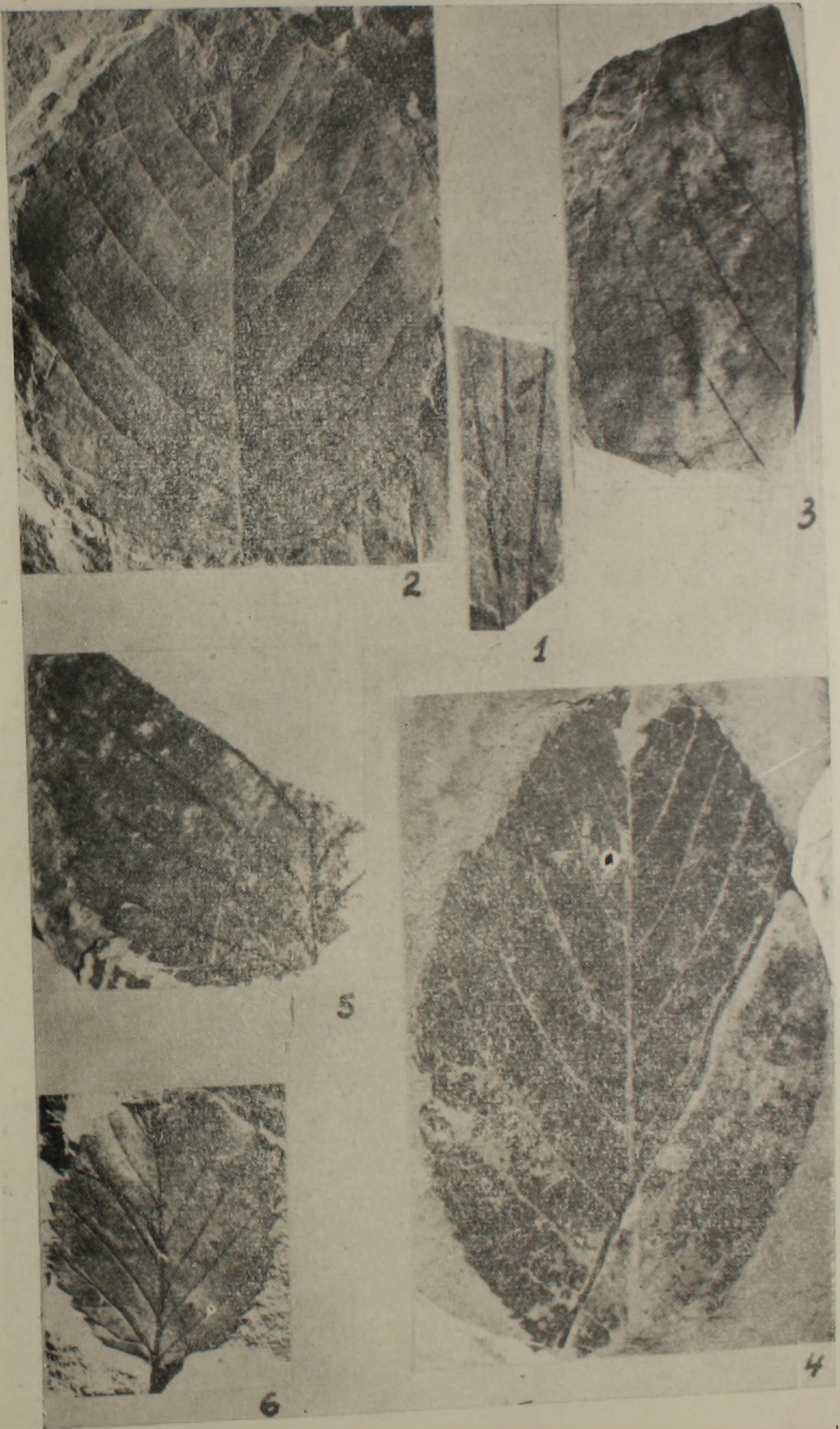
Ա մ փ ո փ ո լ մ

Եթե հատկանիշները կազմում են շարք, ոչ մի առարկա չունի 2 հատկանիշ այդ շարքից, $\sum n_i n_j$ անգամ առարկաները կտարբերվեն իրարից: Առարկաների համեմատության քաշը $= \frac{2 \sum n_i n_j}{m}$, որտեղ $m =$ հատկանիշների, իսկ $n =$ առարկաների թվին: Որոշիչների համար մի քանի շարք հատկանիշների քաշը գումարելը սխալ է իրենց կոռելացիայի պատճառով: Պետք է շարքերի

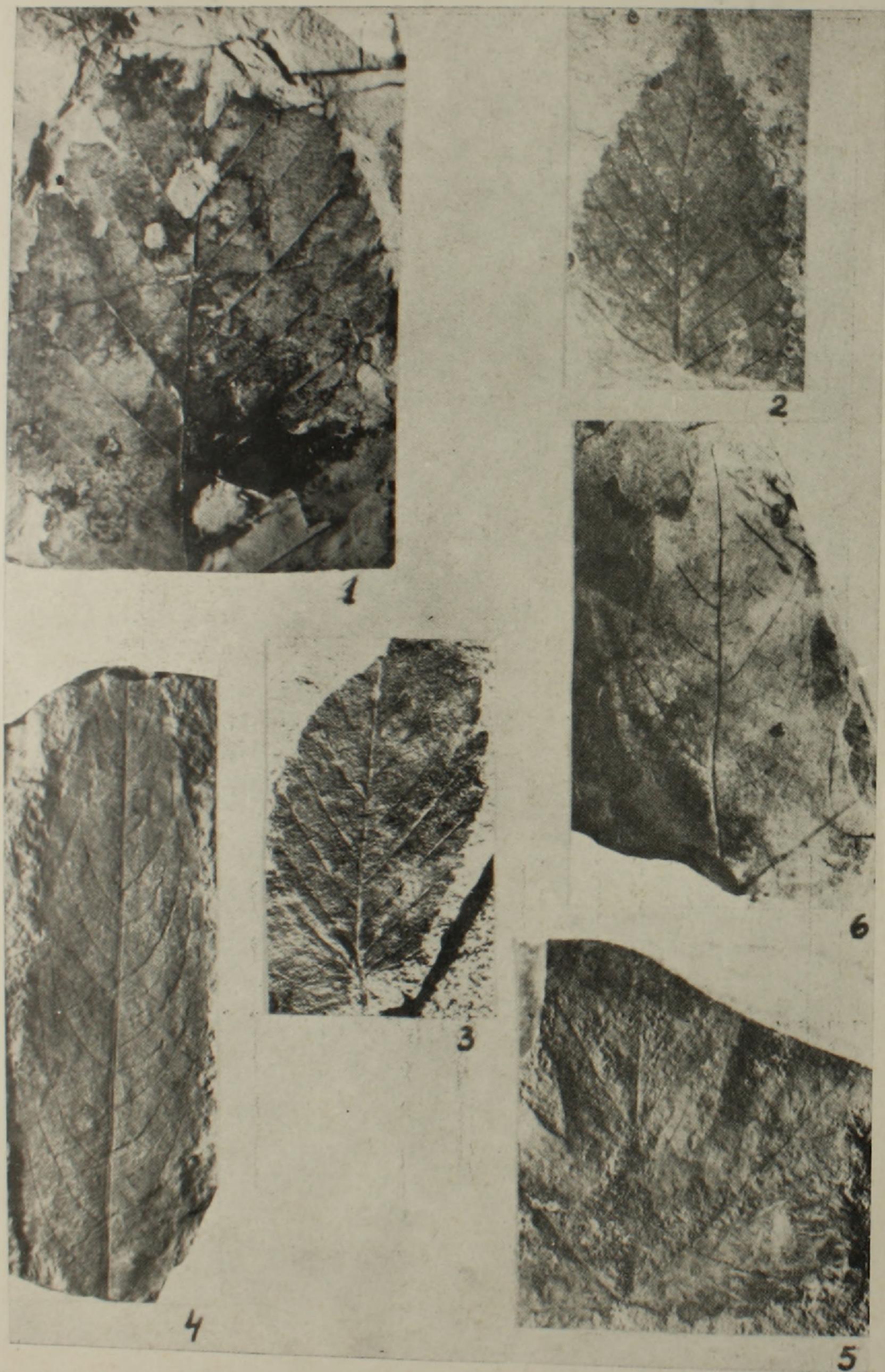
չկրկնվող կոմբինացիան հաշվել մի հատկանիշ և դասակարգման համար բոլոր նմանության գնահատականներից հանել ամենափոքրը:

Л И Т Е Р А Т У Р А

1. Барсегян А., Гамбарян П., Армаганян Н. Уч. зап. ЕГУ, 3, 1971.
2. Бейли Н. Статистические методы в биологии. М., 1962.
3. Вайнштейн Б. А. Журн. общей биол., 29, 2, 1968.
4. Вайнштейн Б. А. Журн. общей биол., 33, 2, 1972.
5. Василевич В. И. Статистические методы в геоботанике. Л., 1969.
6. Гамбарян П. П. Изв. АН АрмССР (биол. науки), 18, 8, 1964.
7. Гамбарян П. П. Биологический журнал Армении, 21, 10, 1968.
8. Гамбарян П. П. Биологический журнал Армении, 22, 10, 1969.
9. Гамбарян П. П. Биологический журнал Армении, 24, 4, 1971.
10. Лобанов А. Л. Энтомолог. обозр. 51, 3, 1972.
11. Расницын А. П. Журн. общей биол., 33, 1, 1972.
12. Смирнов Е. С. Таксономический анализ. М., 1969.
13. Тамарин П. В. Журн. общ. биол., 32, 3, 1971.
14. Тамарин П. В. Журн. общ. биол., 32, 4, 1971.
15. Тамарин П. В. Журн. общ. биол., 33, 5, 1972.
16. Шмидт В. М. Ботанич. журн., 55, 3, 1970.
17. Sokal R. R. and Sneath P. H. A. Principles of numerical taxonomy. S. Fr. et London, 1963.



1 — *Pinus* sp., обр. 82 2 — *Alnus barbata* C. A. Mey., обр. 1. 3, 4 — *A. subcordata* C. A. Mey., обр. 10, 26. 5 — *Corylus* cf. *avellana* L. обр. 61. 6. *Sorbus* cf. *graeca* (Spach) Schauer, обр. 62.



1 — *Betula macrophylla* Heer., обр. 100. 2, 3 — *Carpinus grandis* Ung., обр. 30, 31. 4 — *Salix varians* Goepp., обр. 12. 5 — *Acer hyrcanum* Fisch. et C. A. Mey., обр. 76. 6 — *A. platanoides* L., обр. 37.