Б О Т А Н И Ч Е С К А Я   И Н Ф О Р М А Т И К А

УДК 581.9 : 58.08 : 51

О СВЯЗИ МЕЖДУ СРЕДНИМИ ЗНАЧЕНИЯМИ ДВУХ МЕР ВКЛЮЧЕНИЯ И МЕРАМИ СХОДСТВА.

© Б.И. Сёмкин

Тихоокеанский институт географии ДВО РАН, Владивосток

Рассмотрено два подхода к определению мер сходства. При первом походе мера сходства двух дескриптивных множеств определяется как среднее степенное значение двух мер включения, а при втором – как отношение меры пересечения двух множеств к среднему степенному значению мер сравниваемых множеств.

Рассматриваемые подходы определяют одну и ту же совокупность мер сходства. Приводится иллюстративный пример по сравнительному анализу списков видов сосудистых растений заповедников Приамурья.

Ключевые слова: мера включения, мера сходства, расстояние Юрцева, дескриптивное множество, средняя степенная.

Введение

Меры сходства и различия, включения и невключения широко используются в сравнительной флористике и геоботанике для классификации и ординации флор и сообществ, а также в исследованиях их структурной организации (Галанин, 1991; Осипов, 1992; Кафанов, Жуков, 1993; Ахтямов, 2001; Беликович 2001а,б; Кожевников, Кожевникова, 2004; Петропавловский, 2004; Крестов, 2006). С введением в практику сравнительного анализа флор и сообществ мер включения (Сёмкин, Комарова, 1977; Юрцев, Сёмкин, 1980; Сёмкин, 1987) возникла проблема их связи с мерами сходства (Сёмкин, Комарова, 1985).

Оказалось, что меры сходства двух дескриптивных множеств можно определить посредством различных средних значений между двумя мерами включения (Сёмкин, 1973; 1979; Сёмкин, Комарова, 1977; Юрцев, Сёмкин, 1980) или как частное меры пересечения двух дескриптивных множеств и средних значений мер двух сравниваемых дескриптивных множеств (Песенко, 1982; Раушенбах, 1985; Малышев, 1999).

Подробно рассмотрим два этих подхода к определению мер сходства двух дескриптивных множеств с целью их сопоставления и выбора из них наиболее перспективного для развития теории мер сходства, а также для практического использования. В дальнейшем будем пользоваться простыми общепринятыми теоретико-множественными обозначениями мер сходства и включения (Василевич, 1969; Юрцев, Сёмкин, 1980; Сёмкин, 2007).

1. Первый подход к определению мер сходства

Пусть $a = n(A)$, и $c = n(A \cap B)$, где а – число элементов в множестве А; b – число элементов в множестве В; с – число общих элементов множеств А и В. Тогда две меры включения $K_0(A;B)$ и $K_0(B;A)$ соответственно равны (Сёмкин, 1973; 1979; 2007; Сёмкин, Комарова, 1977; Юрцев, Сёмкин, 1980):

$$K_0(A;B) = \frac{n(A \cap B)}{n(B)} = \frac{c}{b}$$

$$K_0(B;A) = \frac{n(A \cap B)}{n(A)} = \frac{c}{a}$$

Семейство мер сходства двух дескриптивных множеств определяется с помощью средней степенной порядка η значений $K_0(A;B)$ и $K_0(B;A)$ следующим образом (Сёмкин, 1979; 2007; Юрцев, Сёмкин, 1980):

$$K_{0;\eta}(A,B) = \left(\frac{K_0^{\eta}(A;B) + K_0^{\eta}(B;A)}{2} \right)^{1/\eta}$$

, где ; $-\infty < \eta < +\infty$

$$K_{0;0}(A,B) = \sqrt{K_0(A;B) \cdot K_0(B;A)}$$

$$K_{0;-\infty}(A,B) = \min\left(K_0(A;B), K_0(B;A)\right)$$

$$K_{0;+\infty}(A,B) = \max\left(K_0(A;B), K_0(B;A)\right)$$

Как было ранее показано (Сёмкин, 2007), мера $K_{0;+\infty}(A,B)$ является мерой квазисходства, т.е. мера не удовлетворяет одной из аксиом мер сходства. После простых преобразований меру $K_{0;\eta}(A,B)$ можно записать в виде:

$$K_{0;\eta}(A,B) = c\left[\frac{1}{2}\left(\frac{1}{a^{\eta}} + \frac{1}{b^{\eta}}\right)\right]^{1/\eta}$$

Известно (Беккенбах, Беллман, 1965), что $K_{0;\eta}(A,B)$ является монотонной функцией параметра η, т.е.

$$K_{0;-\infty} \le K_{0;-3} \le K_{0;-2} \le K_{0;-1} \le K_{0;0} \le K_{0;1} \le K_{0;2} \le K_{0;3} \le K_{0;+\infty}$$

Приведём некоторые меры сходства, наиболее часто используемые в сравнительном анализе растительного покрова.

$$K_{0;-1} = \frac{2c}{a+b} = \frac{2}{\left(\frac{c}{a}\right)^{-1} + \left(\frac{c}{b}\right)^{-1}}$$ (мера сходства Сёренсена, среднее гармоническое двух мер включения);

$$K_{0;0} = \frac{c}{\sqrt{ab}} = \sqrt{\frac{c}{a}\cdot\frac{c}{b}}$$ (мера сходства Очиаи, среднее геометрическое двух мер включения);

$$K_{0;1} = \frac{1}{2}\left(\frac{c}{a} + \frac{c}{b}\right)$$ (мера сходства Кульчинского, среднее арифметическое двух мер включения);

$$K_{0;-\infty} = \frac{c}{\max(a,b)} = \min\left(\frac{c}{a}, \frac{c}{b}\right)$$ (мера сходства Браун-Бланке, средняя величина как минимальное значение двух мер включения);

$$K_{0;-\infty} = \frac{c}{\max(a,b)} = \min\left(\frac{c}{a}, \frac{c}{b}\right)$$ (мера квазисходства Симпсона, средняя величина как максимальное значение двух мер включения);

Среди указанных мер сходства отсутствует коэффициент Жаккара, так как мы использовали при осреднении только меры включения $K_0(A;B)$ и $K_0(B;A)$.

Если взять семейство мер включения упорядоченного параметром τ (Сёмкин, 1979) $K_{\tau}(A;B)$ и $K_{\tau}(B;A)$ и определить среднее гармоническое значение этих мер включения, то получим следующее семейство мер сходства также упорядоченное параметром τ (Сёмкин, 1973):

$$\frac{1}{K_{\tau;-1}(A,B)} = \frac{1}{2}\left[\frac{1}{K_{\tau}(A;B)} + \frac{1}{K_{\tau}(B;A)}\right],$$

где $K_{\tau}(A;B) = \dfrac{K_0(A;B)}{1+\tau-\tau K_0(A;B)}$,

$$K_{\tau}(B;A) = \frac{K_0(B;A)}{1+\tau-\tau K_0(B;A)}$$

$$-1 < \tau < +\infty$$

$K_{\tau;-1}(A,B)$ – мера сходства. Или в привычных обозначениях:

$$K_{\tau;-1}(A,B) = \frac{2c}{(1+\tau)(a+b) - 2\tau c}$$

При $\tau=1$ получаем меру сходства Жаккара:

$$K_{1;-1}(A,B) = \frac{c}{a+b-c}$$

Мера сходства Жаккара есть среднее гармоническое двух мер включения $K_1(A;B)$ и $K_1(B;A)$ (Сёмкин, Комарова, 1977). Аналогично можно показать, что мера сходства Сокала и Снита $K_{3;-1}(A;B)$ есть среднее гармоническое значение мер включения $K_3(A;B)$ и $K_3(B;A)$ и соответственно равна:

$$K_{3;-1}(A;B) = \frac{c}{2(a+b) - 3c}$$

Расчёты мер сходства посредством определения значения среднего гармонического двух мер включения можно упростить если использовать понятие медианты двух дробей с положительными знаменателями (Хинчин, 1961).

Медианта двух дробей $\frac{a_1}{b_1}$ и $\frac{a_2}{b_2}$ равна:

$$\frac{a_1}{b_1} \oplus \frac{a_2}{b_2} = \frac{a_1 + a_2}{b_1 + b_2}$$

где $\oplus$ – операция взятия медианты, b1 >0, b2 >0, a1 >0, a2 >0.

При одинаковых числителях a1 = a2 = a медианта дробей $\frac{a_1}{b_1}$ и $\frac{a_2}{b_2}$ равна среднему гармоническому этих дробей.

Меры включения двух дескриптивных множеств имеют одинаковые значения числителей равные с. Следовательно, операция взятия среднего гармонического двух мер включения сводится к операции взятия медианты этих мер.

Например, мера сходства Сёренсена просто определяется через меры включения:

$$K_{0;-1}(A,B)=\frac{c}{a}\oplus\frac{c}{b}=\frac{c+c}{a+b}=\frac{2c}{a+b}$$

Аналогично определим и меру сходства Жаккара:

$$K_{1;-1}(A,B)=\frac{c}{2a-c}\oplus\frac{c}{2b-c}=\frac{2c}{2a+2b-2c}=\frac{c}{a+b-c}$$

Также определим меру Сокала и Снита:

.

$$K_{3;-1}(A,B)=\frac{c}{4a-3c}\oplus\frac{c}{4b-3c}=\frac{2c}{4(a+b)-6c}=\frac{c}{2(a+b)-3c}$$

Второй подход к определению мер сходства

По этому способу мера сходства $K'_{0;t}(A,B)$ определяется следующим образом:

$$K'_{0;t}(A,B)=\frac{c}{\left[\frac{1}{2}\left(a^t+b^t\right)\right]^{1/t}}$$

где $-\infty\le t\le+\infty$

Особо отметим три значения параметра t (-∞, + ∞, 0), при которых, при раскрытии неопределённости, получим следующие меры:

$$K'_{0;-\infty}(A,B)=\frac{c}{\min(a,b)}.$$

$$K'_{0;+\infty}(A,B)=\frac{c}{\max(a,b)},$$

$$K'_{0;0}(A,B)=\frac{c}{\sqrt{ab}}$$

Следует отметить, что совокупность мер сходства полученных с помощью, как первого, так и второго способа совпадают. Однако первый способ имеет преимущество перед вторым, так как позволяет вычислять по матрице мер включения меры сходства путём различных методов симметризации.

Сравнение подходов

Легко вывести соотношение между мерами: $K_{0;\eta}(A,B)$и:

$$K'_{0;\eta}(A,B)=K_{0;-\eta}(A,B)$$

Следовательно, эти меры совпадают между собой, но имеют различное упорядочивание по параметру η (взаимно обратное).

Пусть, например, $\eta=1$, тогда эти меры сходства соответственно равны $K_{0;-1}=K'_{0;1}$ или $K_{0;-1}=\frac{2c}{a+b}$ – мера сходства Сёренсена (среднее гармоническое двух мер включения $\frac{c}{a}$ и $\frac{c}{b}$),

$$K'_{0;1}=\frac{c}{\left(\frac{a+b}{2}\right)}=\frac{2c}{a+b}$$

также является мерой сходства Сёренсена и равна отношению с к среднему арифметическому двух величин a и b. При η =-1 соответственно получим $K_{0;1}=\frac{1}{2}\left(\frac{c}{a}+\frac{c}{b}\right)$ – мера сходства Кульчинского, равная среднему арифметическому двух мер включения $\frac{c}{a}$ и $\frac{c}{b}$, а

$$K'_{0;-1}=\frac{c}{\left[\frac{1}{2}\left(\frac{1}{a}+\frac{1}{b}\right)\right]^{-1}}=\frac{1}{2}c\left(\frac{1}{a}+\frac{1}{b}\right)$$

– мера сходства, равная отношению с к среднему гармоническому двух величин a и b.

На основании полученного соотношения между мерами сходства $K_{0;\eta}(A;B)$ и $K'_{0;\eta}(A;B)$ легко решается спорный вопрос, поднятый Л.И. Малышевым (1999). Он в частности утверждает, что «Б.И. Сёмкин (1987) ошибочно полагал, что формула Чекановского учитывает сходство по среднему гармоническому отношению, но в действительности это справедливо для формулы Кульчинского (см. ниже), так как формула Чекановского (или Сёренсена) учитывает лишь среднее арифметическое отношение».

Но на самом деле никакой ошибки допущено не было. Как было показано нами, формула Сёренсена (Чекановского) действительно представляет собой с одной стороны среднее гармоническое двух мер включения (Сёмкин, 1987), с другой стороны, отношение числа общих видов к среднему арифметическому численности видов в первой и второй флоре (Малышев, 1999). Аналогично формула Кульчинского (мера сходства) представляет с одной стороны среднее арифметическое двух мер включения (Сёмкин, 1987), а с другой – отношение числа общих видов к среднему гармоническому численностей видов первого и второго списков (Малышев, 1999). Следовательно, в работе Б.И. Сёмкина (1987) ошибки нет.

Как было отмечено выше, совокупность мер сходства $K_{0;\eta}(A;B)$ и $K'_{0;\eta}(A;B)$ ($-\infty\le\eta\le+\infty$) содержит одинаковый набор мер сходства, но по разному упорядоченный. Например, меры сходства, предложенные Ю.А. Песенко (Песенко, 1982 : 140) $K'_{0;-2}(A,B)$ и $K'_{0;-3}(A,B)$ (в наших обозначениях), были известны и равны соответственно $K_{0;2}(A,B)$ и $K_{0;3}(A,B)$ (Сёмкин, 1979).

Равенство мер сходства $K_{0;-\infty}(A,B)$ и $K'_{0;+\infty}(A,B)$, а также мер квазисходства $K_{0;+\infty}(A,B)$ и $K'_{0;-\infty}(A,B)$, доказано в работе (Сёмкин, Комарова, 1977), где приведены следующие соотношения, соответственно:

$$\min\left(\frac{c}{a},\frac{c}{b}\right)=\frac{c}{\max(a,b)};$$

$$\max\left(\frac{c}{a},\frac{c}{b}\right)=\frac{c}{\min(a,b)}$$

Обоснование метода осреднения двух мер включения

Меры сходства двух дескриптивных множеств можно определить с помощью метода осреднения двух мер включения (Сёмкин, 1973; 1979; 1987; 2007; Сёмкин, Юрцев, 1980).

Наиболее общий вид, так называемых ассоциативных средних, был введён в 1930-1931 гг. А.Н. Колмогоровым, Нагуно и де Фонетти (Джини, 1970, С. 150; Орлов, 1977, С. 18):

$$\mathrm{g_p}(X,F)=F^{-1}\left(\frac{1}{p}\sum_{i=1}^{p}F(x_i)\right)$$

где F (Z) – строго монотонная функция; F-1 (Z) – обратная предыдущей функция; P>0 – целое положительное число; $X=(x_1, x_2,..., x_P)$, $x_i \geq 0$ $(i=1, 2,..., P)$ – неотрицательные действительные числа.

Установлено (Орлов, 1977), что $\mathrm{g_p}(X,F_1)=\mathrm{g_p}(X,F_2)$ тогда и только тогда когда $\mathrm{F_2}(Z)=a\mathrm{F_1}(Z)+d$ при некоторых a и d.

При измерении признаков в шкале отношений (или в частном случае, в абсолютной шкале) математически корректно пользоваться функцией $\mathrm{F}(Z)=Z^{\eta}$, η≠0; *ln* Z или

$$\mathrm{g_p}(X)=\left(\frac{1}{p}\sum_{i=1}^{p}x_i^{\eta}\right)^{1/\eta},\text{при } \eta\neq 0,\ \mathrm{g_p}(X)=\left(\prod_{i=1}^{p}x_i\right)^{1/p}$$

то есть средней степенной (Орлов, 1977; 1981; Кузьмин, Орлов, 1977; Раушенбах, 1987).

Определение симметричных мер посредством нахождения различных средних величин двух несимметричных мер предложено Б.И. Сёмкиным (1973; 1979), при этом в качестве средней величины использовались средние степенные. Так, с помощью взятия средней гармонической для двух дескриптивных множеств определена относительная мера сходства, обобщающая меру Сёренсена на случай весовых множеств (Сёмкин, 1973). Осредняя две меры включения с помощью средней степенной, получаем симметричные меры сходства, упорядоченные двумя параметрами – τ и η (Сёмкин, 1979).

Приведённый подход осреднения мер включения использовался для симметризации матрицы несимметричных мер включения (Сёмкин, 1987; Сёмкин, Комарова, 1985). В матрице мер включения для элементов, симметричных относительно главной диагонали, определяется некоторая их средняя величина и в результате операции получаем симметричную матрицу мер сходства. Наиболее часто используется средние величины – средняя гармоническая, арифметическая, геометрическая, посредством которых определяется соответственно меры сходства Сёренсена, Кульчинского и Очиаи. Иногда в качестве средних величин используется операция нахождения минимального и максимального значений двух мер включения (Сёмкин, Комарова, 1985).

Наиболее адекватной мерой сходства для измерения отношений сходства между двумя дескриптивными множествами является мера Браун-Бланке и соответствующая ей мера различия Юрцева.

Остановимся на обосновании наиболее корректного метода осреднения мер включения. Определим четыре типа отношений:

1) Отношение несходства при пороге δ. Дескриптивные множества А и В несходны при пороге δ, если $K(A;B)<\delta$ и $K(B;A)<\delta$.

2) Отношение сходства при пороге δ. Дескриптивные множества А и В сходны при пороге δ, если $K(A;B)\geq\delta$ и $K(B;A)\geq\delta$.

3) Отношение включения А в В при пороге δ. Дескриптивное множество А включается в дескриптивное множество В при пороге δ, если $K(B;A)\geq\delta$ и $K(A;B)<\delta$.

4) Отношение включения В в А при пороге δ. Дескриптивное множество В включается в дескриптивное множество А при пороге δ, если $K(B;A)<\delta$ и $K(A;B)\geq\delta$.

Отношение сходства также можно определить с помощью порога δ и меры сходства $K(A,B)$. Дескриптивные множества А и В сходны, если мера сходства $K(A,B)$ не меньше порога δ, т.е. $K(A,B)\geq\delta$. Будем считать, что меры включения $K(A;B)$ и $K(B;A)$ взаимно согласованы (двухсторонне согласованы) с мерой сходства $K(A,B)$, если справедливы следующие соотношения между ними: «мера сходства $K(A,B)\geq\delta$ тогда и только тогда, когда $K(A;B)\geq\delta$ и $K(B;A)\geq\delta$ ».

Таблица 1

Матрица пересечений заповедников Приамурья, рассчитанная по аборигенной фракции видов сосудистых растений (по: Кожевников, Близнюк, 2005)

	1	*2*	*3*	*4*	*5*	*6*
1	**502**	295	211	222	218	211
2	295	**601**	391	353	387	337
3	211	391	**788**	465	635	379
4	222	353	465	**618**	541	342
5	218	387	635	541	**853**	375
6	211	337	379	342	375	**491**

Примечание. Здесь и далее в голове таблиц заповедники Приамурья: 1 – Буреинский; 2 – Зейский; 3 – Хинганский; 4 – Комсомольский; 5 – Большехехцирский; 6 – Норский

Таблица 2

Матрица мер включения $K_0(A;B)$ и $K_0(B;A)$, в %, рассчитанная на основе матрицы пересечений (табл. 1)

	1	*2*	*3*	*4*	*5*	*6*
1	**100**	59	42	44	43	42
2	49	**100**	65	59	64	56
3	27	50	**100**	59	81	48
4	36	57	75	**100**	88	55
5	26	45	74	63	**100**	44
6	43	69	77	70	76	100

Таблица 3

Матрица мер включения $K_1(A;B)$ и $K_1(B;A)$

→	*1*	*2*	*3*	*4*	*5*	*6*
1	**100**	42	27	28	27	27
2	32	**100**	48	42	47	39
3	16	33	**100**	12	68	32
4	22	40	60	**100**	79	38
5	15	29	59	46	**100**	28
6	27	53	63	54	61	100

Указанному условию удовлетворяет только мера сходства Браун-Бланке и эквивалентные ей меры. Действительно, если $K_{0;-\infty}(A,B) = \min\left(K_0(B;A), K_0(A;B)\right) \ge \delta$, то в этом случае должно быть $K_0(A;B) \ge \delta$ и $K_0(B;A) \ge \delta$. Обратно, если $K_0(A;B) \ge \delta$ и $K_0(B;A) \ge \delta$, то $K_0(B;A) \ge \delta$.

Справедливо также двойственное соотношение для меры различия $F(A,B) = 1 - K(A,B)$: «мера различия $F(A,B) \le 1 - \delta$ тогда и только тогда, когда мера невключения $F(A;B) = 1 - K(A;B) \le 1 - \delta$ и мера невключения $F(B;A) = 1 - K(B;A) \le 1 - \delta$ ».

Указанным условиям удовлетворяет только одна мера различия (расстояние Юрцева), а также эквивалентные ей меры.

Примеры симметризации матриц мер включения

Рассмотрим пример по установлению флористических связей Норского заповедника с заповедниками Приамурья, по данным А.Е. Кожевникова и Т.Н. Близнюк (2004).

Для установления связей между флорами используется только аборигенная фракция видов сосудистых растений каждого заповедника. Основой расчётов мер включения и сходства является матрица абсолютных мер сходства (матрица пересечения) (табл. 1), содержащая информацию о количестве видов в каждой флоре (диагональ матрицы) и о количестве общих видов для каждой пары сравниваемых флор (Кожевников, Близнюк, 2004). Следует отметить, что в случае отсутствия в публикациях списков видов флор основным «документом», позволяющим воспроизвести результаты расчётов матриц включения, сходства и других, является именно матрица пересечения.

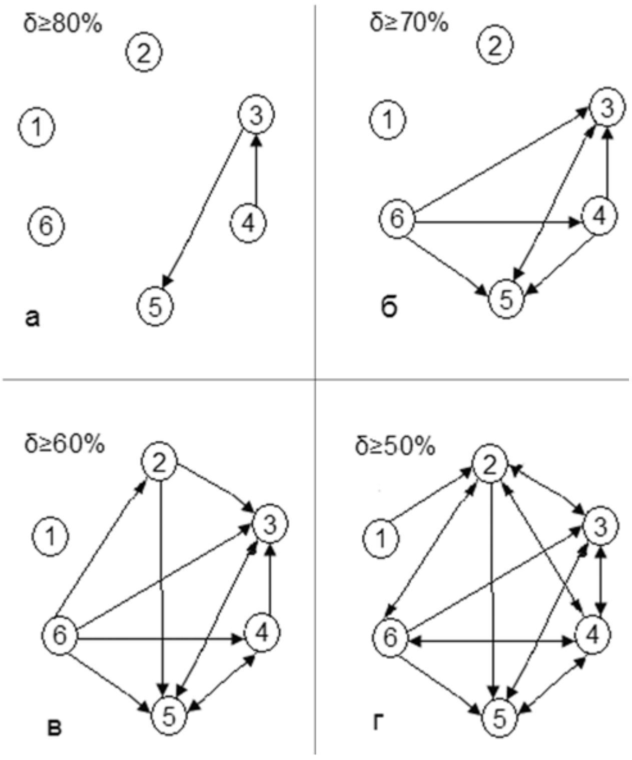

Рис. 1. Графы включения-сходства флор 6 заповедников Приамурья (а – при пороге δ≥80%; б – при пороге δ≥70%; в – при пороге δ≥60%; г – при пороге δ≥50%;)

Диагональ матрицы пересечений (табл. 1) содержит значительно разновеликие числовые значения. Например, число видов в Норском заповеднике равно 491, а в Большехекцирском – 853. Так как численность видов разновелика в каждом заповеднике, мы будем проводить анализ несимметричных отношений. Рассмотрим три метода симметризации матриц мер включения.

МЕРЫ ВКЛЮЧЕНИЯ

По матрице мер пересечений (табл. 1) рассчитаем матрицу мер включения (табл. 2). Для этого необходимо каждый элемента строки матрицы пересечений разделить на соответствующий диагональный элемент, а результат выразить в процентах. Например, диагональный элемент первой строки равен 502, на этот элемент делим каждый элемент первой строки матрицы пересечений и получаем первую строку матрицы мер включений. Повторяя аналогичные расчёты для 2-ой, 3-ей и других строк матрицы пересечений, получим матрицу мер включения (табл. 2). В левом углу матрицы включения приводится стрелка, указывающая направление включения.

Проведём анализ связей по матрице мер включения (табл. 2). При порогах δ≥80%, δ≥70%, δ≥60% и δ≥50% получаем графы включения-сходства (рис. 1). Граф включения показывает, что наиболее сильные флористические несимметричные связи при пороге δ≥80% (рис. 1а) обнаружены между флорами трёх заповедников: флора Большехекцирского заповедника (5) включает как часть флору Хинганского заповедника (3), а последний включает как часть флору Комсомольского заповедника (4).

При рассмотрении более слабых связей, при пороге δ≥70% (рис. 1б), устанавливается несколько отличная структура. Флоры Большехекцирского (5) и Хинганского (3) заповедников сходны и в них включаются как части флоры Комсомольского (4) и Норского (6) заповедников. Норский заповедник также включается и в Комсомольский. Заповедники Буреинский (1) и Зейский (2) не включаются в другие.

Рассмотрим ещё более слабые связи при δ≥60% (рис. 1в). Наблюдаем сходство флор двух

Таблица 4

Матрица мер сходства (Сёренсена), в %, полученная с помощью симметризации матрицы мер включения (табл. 2)

	1	*2*	*3*	*4*	*5*	*6*
1	**100**	54	34	40	34	42
2	54	**100**	58	58	54	62
3	34	58	**100**	67	78	62
4	40	58	67	**100**	76	62
5	34	54	78	76	**100**	60
6	42	62	62	62	60	100

заповедников (Большехекцирский, Хинганский), также добавляется односторонняя связь Норского заповедника (6) с Зейским (2). Флора Зейского заповедника, в свою очередь, включается во флоры Большехекцирского и Хинганского. Флора Буреинского заповедника (1) при данном пороге в флоры других заповедников не включается.

Наконец, учитывая достаточно слабые связи (рис. 1г) при пороге δ≥50%, все заповедники (исключение составляют Буреинский и Зейский) связаны взаимными связями сходства. Флора Буреинского заповедника включается только во флору Зейского заповедника. Односторонняя связь наблюдается у Зейского заповедника с Большехекцирским.

В целом, следует отметить, что все приведённые заповедники Приамурья весьма сходны между собой, исключая только Буреинский заповедник.

Проведём расчёты эквивалентных мер включения $K_1(A;B)$ и $K_1(B;\text{A})$ по следующим формулам:

$$K_1(A;B)=\frac{K_0(A;B)}{2-K_0(A;B)},\quad K_1(\text{B};\text{A})=\frac{K_0(\text{B};\text{A})}{2-K_0(\text{B};\text{A})}$$

В результате получим матрицу мер включения

$$K_1(A;B)=\frac{c}{2b-c},\quad K_1(\text{B};\text{A})=\frac{c}{2a-c}$$, (табл. 3).

Проведём анализ связей между флорами заповедников Приамурья по матрице мер включения $K_1(A;B)$ и $K_1(B;\text{A})$.

Если произвести пересчёт порогов δ по формуле $\delta_1=\dfrac{\delta\cdot 100\%}{200-\delta}$, то получим следующие значения порогов $\delta_I\geq 67\%$, $\delta_I\geq 54\%$, $\delta_I\geq 43\%$ и $\delta_I\geq 33\%$. При указанных порогах структура связей между рассматриваемыми флорами, полученная на основе матрицы мер включения $K_1(A;B)$ и $K_1(B;\text{A})$, остаётся прежней, как на рис. 1. В этом заключается эквивалентность мер включения K_0 и K_1.

МЕРА СХОДСТВА СЁРЕНСЕНА

Проведём осреднение элементов матрицы мер включения (табл. 2) симметричных относительно диагонали, используя среднюю гармоническую двух величин. В результате получим симметризованную матрицу мер сходства Сёренсена (табл. 4).

Две меры включения несут больше информации о связях, чем одна симметричная мера сходства. Если в матрице мер включения (табл. 2) были такие сильные связи как $K(5;4)=88\%$, $K(5;3)=81\%$, то наибольшие связи в матрице сходства (табл. 4) равны $K(5;4)=76\%$, $K(5;3)=78\%$.

Рассмотрим структуру связей между флорами заповедников Приамурья на основе матрицы мер сходства Сёренсена (табл. 4) и матрицы мер включения K_0(табл. 2). При пороге δ≥59% получаем граф включения-сходства (рис. 2). Как видно из этого графа, заповедники Большехехцирский, Хинганский, Комсомольский и Норский сходны между собой. Флора Буреинского заповедника включается только в флору Зейского, а флора последнего в свою очередь включается в флору Большехехцирского, Хинганского и Комсомольского.

МЕРА СХОДСТВА ЖАККАРА

Проведём осреднение элементов матрицы мер включения K_1 симметричных относительно диагонали, используя среднюю гармоническую двух величин. В результате получим симметризованную матрицу мер сходства Жаккара (табл. 5, над диагональю) и рассчитаем двойственную ей матрицу мер различия (табл. 5, под диагональю).

При пороге $\delta_I\geq 42\%$ по матрице сходства Жаккара и матрице мер включения K_1 можно построить граф включения-сходства, идентичный ранее построенному графу (рис. 2). В этом примере построим оптимальное дерево (дендрит) (рис. 3) на основе матрицы мер различия Жаккара (табл. 5,

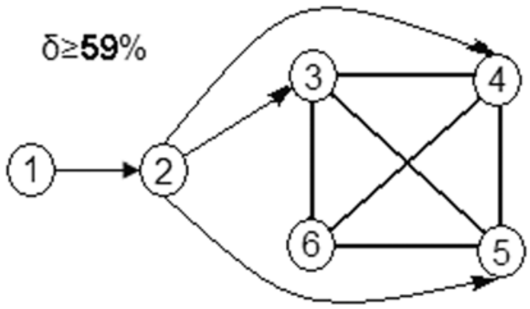

Рис. 2. Граф включения-сходства, построенный по матрице мер сходства Сёренсена и соответствующей матрицы мер включения (табл. 2) при пороге δ≥59%.

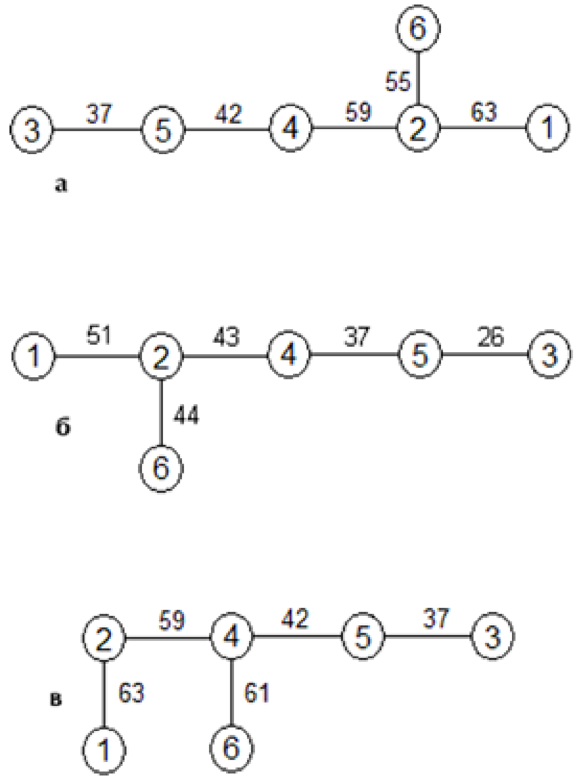

Рис. 3. Оптимальное дерево (дендрит), построенное на основе матрицы мер различия Жаккара (а, б) и построенное на основе матрицы расстояний Юрцева (в).

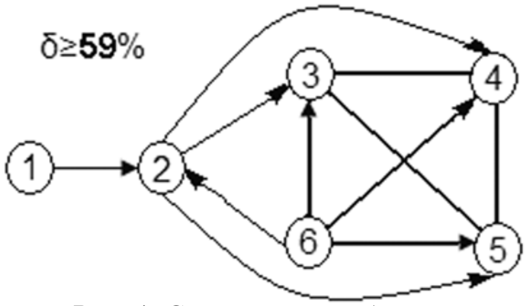

Рис. 4. Смешанный граф включения-сходства, построенный по матрице мер включения K0 (табл. 2) и мер сходства Браун-Бланке (табл. 6) при пороге δ≥59%.

под диагональю), которая является двойственной матрице сходства Жаккара.

Задача нахождения кратчайшего дерева оказалась неоднозначной – получилось два оптимальных дерева. Флора Норского заповедника в равной степени сходна как с флорой Зейского заповедника, так и с флорой Комсомольского заповедника. Наименьшие различия флор имеют Большехекцирский и Хинганский заповедники, флора Комсомольского близка к флоре Большехекцирского заповедника. По-видимому, флора Норского заповедника занимает промежуточное положение (в равной степени тяготеет как к Комсомольскому, так и к Зейскому). Это же было отмечено и в работе А.Е. Кожевникова и Т.Н. Близнюк (2004).

МЕРА СХОДСТВА БРАУН-БЛАНКЕ И РАССТОЯНИЕ ЮРЦЕВА

Как было показано нами, для определения отношений сходства между дескриптивными множествами можно использовать ещё и меру сходства Браун-Бланке, а также двойственную ей меру различия (расстояние) Юрцева. Проведём симметризацию матрицы мер включения K_0 (табл. 2) путём взятия минимальных элементов для симметричных элементов относительно главной диагонали. В результате получим матрицу мер сходства Браун-Бланке (табл. 6, над диагональю).

По матрице сходства Браун-Бланке при пороге δ≥59% построим граф включения-сходства (рис. 4). В результате получаем достаточно простую схему флористических связей между флорами заповедников Приамурья. Большехехцирский, Хинганский и Комсомольский заповедники сходны при этом пороге, в них включается флора Норского заповедника, а последняя во флору Зейского. Флора Зейского, в свою очередь, включается во флоры Комсомольского, Хинганского и Большехекцирского заповедников. Флора Буреинского заповедника включается только в флору Зейского.

По формуле $100-\max\left(K_0(A;B),K_0(B;A)\right)$ определим расстояние Юрцева (Сёмкин, 2007) и представим попарные расстояния в виде матрицы (табл. 6, под диагональю). По полученной матрице расстояния построим оптимальное дерево (дерево кратчайших расстояний) (Рис. 3).

Как видим из исследования дендрита, флора Норского заповедника ближе всех к флоре Зейского, его же флора близка к флоре Комсомольского. В целом дендрит (рис. 3) наглядно отражает упорядочивание флор 6 заповедников по флористическому сходству и показывает по-

Таблица 5

Матрица мер сходства (Жаккара), полученная с помощью симметризации матрицы мер включения (табл. 4), в % (над диагональю) и матрица мер различия Жаккара $F_1 = (100 - K_1)$ в % (под диагональю)

	1	*2*	*3*	*4*	*5*	*6*
1	×	37	20	25	19	27
2	63	×	39	41	36	45
3	80	61	×	49	63	42
4	75	59	51	×	58	45
5	81	64	37	42	×	39
6	73	55	58	55	61	×

Таблица 6

Матрица мер сходства (Браун-Бланке), полученная с помощью симметризации матрицы мер включения (табл. 2) в % (над диагональю) и матрица мер расстояния (Юрцева) в % (под диагональю)

	1	*2*	*3*	*4*	*5*	*6*
1	×	49	27	36	26	42
2	51	×	50	57	45	56
3	73	50	×	59	74	48
4	64	43	41	×	63	55
5	74	55	26	37	×	44
6	58	44	52	45	56	×

ложение флоры Норского заповедника среди других исследуемых флор.

Заключение

Рассмотренные подходы к определению мер сходства с помощью осреденения определяют одни и те же меры сходства, но различно упорядочивают их совокупность. С теоретической и практической точек зрения предпочтение следует отдавать первому подходу, при котором осредняются величины мер включения. Из рассмотренных мер сходства и различия только мера сходства Браун-Бланке и двойственная ей мера различия (расстояние) Юрцева удовлетворяют условию согласования меры сходства (различия) с соответствующими мерами включения, что позволяет корректно представлять отношения включения-сходства (или двойственные им отношения) и графически изображать в виде смешанных графов, имеющих как дуги для характеристики включения, так и рёбра, характеризующие отношения сходства.

При упорядочивании флор заповедников Приамурья наиболее содержательно интерпретируемые результаты были нами получены при использовании меры сходства Браун-Бланке и меры различия (расстояния) Юрцева.

Литература

Ахтямов М.Х. Ценотаксономия прирусловых ивовых и ивово-тополёвых и уремных лесов поймы реки Амур. – Владивосток: Дальнаука, 2001. – 138 с.

Беккенбах Э., Беллман Р. Неравенства. – М.: Мир, 1965. – 276 с.

Беликович А.В. Ландшафтно-флористическая неоднородность растительного покрова (на примере модельных районов Северо-Востока России). – Владивосток: БПИ ДВО РАН, 2001. – 248 с.

Беликович А.В. Растительный покров северной части Корякского нагорья. . – Владивосток: Дальнаука, 2001. – 420 с.

Василевич В.И. Статистические методы в геоботанике. Л., Наука, 1969. 232 с.

Галанин А.В. Флора и ландшафтно-экологическая структура растительного покрова. – Владивосток: ДВО АН СССР, 1991. – 272 с.

Джини К. Средние величины. – М. Статистика. 1970. – 448 с.

Кафанов А.И., Жуков В.Е. Прибрежное сообщество водорослей-макрофитов залива Посьета (Японское море): Сезонная изменчивость и пространственная структура. – Владивосток: Дальнаука, 1993. – 156 с.

Кожевников А.Е., Близнюк Т.Н. Основные особенности растительного покрова и флористические связи Норского заповедника в Приамурье // Ботанические исследования в Приамурье и на сопредельных территориях: Матер. регионального совещания, Благовещенск, 24-26 мая 2004 г. – Благовещенск: АФБСИ ДВО РАН. 2005. – С. 33–39.

Кожевников А.Е., Кожевникова З.В. Эффективность охраны сосудистых растений Приморья и Приамурья на заповедных территориях // Вестник ДВО РАН. 2004. – № 4. – С. 8–22.

Крестов П.В. Растительный покров и фитогеографические линии Северной Пацифики // Автореф. дисс.... д-ра биол. наук. – Владивосток, 2006. – 42 с.

Кузьмин В.Б., Орлов А.И. О средних величинах, сравнение которых инвариантно относительно допустимых преобразований шкалы // Статистические методы анализа экспертных оценок / Учёные записки по статистике. – Т.29. – М.: Наука, 1977. – С. 220–227.

Малышев Л.И. Основы флористического районирования // Бот. журн. 1999. – Т. 84. № 1. – С. 3–14.

Орлов А.И. Проблема устойчивости и обоснованности решений в теории экспертных оценок // Статистические методы анализа экспертных оценок / Учёные записки по статистике. – Т.29. – М.: Наука, 1977. – С. 220–227.

Орлов А.И. Связь между средними величинами и допустимыми преобразованиями шкалы // Мат. заметки. 1981. – Т.30, №4. – С. 7–30.

Осипов С.В. Изучение строения растительного покрова на основе сравнения соседних участков // Бот. журн. 1992. – Т. 77. № 8. – С. 127–135.

Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. – М.: Наука, 1982. – 287 с.

Петропавловский Б.С. Леса Приморского края: (Эколого-географический анализ). – Владивосток: Дальнаука, 2004. – 317 с.

Раушенбах Г.В. Меры близости и сходства // Анализ нечисловой информации в социологических исследованиях. – М.: Наука, 1985. – С. 169–203.

Сёмкин Б.И. Дескриптивные множества и их приложения // Исследование систем. Т. 1. Анализ сложных систем. – Владивосток. ДВНЦ АН СССР, 1973. – С. 83–94.

Сёмкин Б.И. К методике анализа разновеликих множеств в сравнительной флористике // Комаровские чтения. – Вып. LVI. 2008. – С. 179–195.

Сёмкин Б.И. Количественные показатели для оценки односторонних флористических связей, предложенных Б.А. Юрцевым // Бот. журн. 2007. – Т. 92. № 4. – С. 114–127.

Сёмкин Б.И. Теоретико-графовые методы в сравнительной флористике // Теоретические и методические проблемы сравнительной флористики: Матер. II рабочего совещания по сравнительной флористике. Неринга, 1983. – Л.: Наука. 1987. – С. 149–163.

Сёмкин Б.И. Эквивалентность мер близости и иерархическая классификация многомерных данных // Иерархические классификационные построения в географической экологии и систематике. – Владивосток. ДВНЦ АН СССР, 1979. – С. 97–112.

Сёмкин Б.И., Комарова Т.А. Анализ фитоценотических описаний с использованием мер включения (на примере растительных сообществ долины реки Амгуэмы на Чукотке) // Бот. журн. 1977. – Т. 62. № 1. – С. 54–63.

Сёмкин Б.И., Комарова Т.А. Использование мер включения при изучении вторичных сукцессий (на примере послепожарных сообществ Сихотэ-Алиня) // Бот. журн. 1985. – Т. 70. № 1. – С. 89–97.

Хинчин А.Я. Цепные дроби. – М.: Физматлит, 1961. – 112 с.

Юрцев Б.А., Сёмкин Б.И. Изучение конкретных и парциальных флор с помощью математических методов // Бот. журн. 1980. – Т. 65. № 12. – С. 1706–1718.

On the relation between mean values of two measures of inclusion and measures of similarity

B.I. Semkin

Pacific Institute of Geography FEB of RAS, Vladivostok

Key words: measure of inclusion, measure of similarity, Yurtsev's distance, descriptive set, power mean.

Two approaches to determination of the similarity measures are considered. In the first approach, the similarity measure of two descriptive sets is determined as the average power value of two inclusion measures while the second one defines the similarity measure as the ratio of the intersection measure of two sets to the average power value of measure of the sets compared.

The approaches considered determine the same set of similarity measures. The illustrative example of the comparative analysis is given for the list of species of vascular plants for the state reserves of the Amur Basin Region.

Tabl. 6. Ill. 4. Bibl. 28.