ОБНАРУЖЕНИЕ ЭМПИРИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ С ПОМОЩЬЮ ЭВМ
(Вычислительные системы)
1984 год Выпуск 102

УДК 519.8-519.5

К ТЕОРИИ МЕР СВЯЗИ

Ю.А.Воронин, Н.Г.Горелова

Необходимость введения некоторой меры для оценки степени связи между двумя свойствами возникает в задачах анализа данных при выборе информативной совокупности свойств или при прогнозировании значений одного свойства через значения другого.

При решении конкретной задачи, как правило, имеются, с одной стороны некоторые априорные предположения о поведении меры на исследуемом множестве свойств и, с другой стороны, экспериментальные данные или материал обучения, когда относительно некоторых пар свойств указаны в порядковой (больше или меньше связаны свойства) или в абсолютной шкале значения меры связи.

Теория мер связи, как нам представляется, в первую очередь призвана:

1) систематизировать известные меры связи и предложить регулярный подход к построению новых;

2) указать необходимые и достаточные условия для выбора меры связи в конкретной ситуации;

3) разработать процедуры выбора и оценки параметров мер связи в соответствии с некоторыми внешними критериями.

Основное внимание в настоящей статье сосредоточено на 1-м вопросе, а именно на представлениях о множестве допустимых мер связи, на том, какими свойствами (возможно, разными в различных ситуациях) должны характеризоваться меры и на построении множества параметрических мер сходства, его классификации и исследовании.

В качестве примера предлагается рассмотреть частный случай построения мер связи между ранжировками. Меры связи между ранжировками имеют и самостоятельное значение, например, при сопоставлении стратиграфических разрезов. Кроме того, любое свойство $f_\alpha \in F$

порождает упорядоченное (направленное или ненаправленное) или неупорядоченное разбиение множества A на классы неотличимости $\{A_i\}_\alpha$. В упорядоченном направленном разбиении каждой паре элементов приписано отношение предпочтения $r^>$, если элементы принад - лежат разным классам, и отношение неотличимости $r^\sim$, если одному. Если в каждом классе неотличимости содержится только по одному элементу, говорим о строгих упорядоченных разбиениях, если есть класс с большим числом элементов - о квазиупорядоченных разбиениях.

Первичным заданием упорядоченного разбиения считаем простое перечисление имен элементов в порядке их предпочтения: $A_i = \langle a_1, a_2, \ldots, (a_k, \ldots, a_1), \ldots, a_N \rangle$. Круглыми скобками объединены элементы одного класса неотличимости, N - число элементов, n - число классов неотличимости. Такое упорядоченное разбиение может быть описано матрицей смежности, графом, а также рангами или номерами. Каждому описанию соответствуют свои понятийная база и аппарат исследования. Вопрос о выборе базового описания при построении теории безусловно является очень важным и требует специального изучения. Номерное описание кажется нам наиболее экономным и удобным, поэтому мы останавливаемся на нем [4, с.4]. Зафиксируем некоторый базовый или алфавитный порядок элементов $a_1, a_2, \ldots, a_N$ и опишем его номерами с I по N, тогда любое упорядоченное разбиение A_i представляем в виде последовательности номеров $P_i = \langle N_{i1}, N_{i2}, \ldots, N_{iN} \rangle$, где N_{ik} - номер или ранг k-го элемента алфавита в A_i.

Вообще говоря, термины: упорядочение, ранжировка, порядок (строгие или квази-) - используются как синонимы. В работе термин ранжировка принят за основной, чтобы подчеркнуть номерное описание, а также в соответствии с ГОСТом [4].

Необходимо оговорить также следующее. По способу получения различаем объективные (например, полученные с помощью приборов) и субъективные или экспертные ранжировки. По всей видимости, имеют в виду разные вещи, когда говорят о связи ранжировок, полученных одним способом и о связи ранжировок, полученных разными способами. Особого рассмотрения заслуживает вопрос о сравнении квазиранжи - ровок с разным числом классов. Пусть имеются разбиения $\{A_i\}_1$ и $\{A_i\}_2$ с числом классов $n_1 < n_2$. Из разбиения $\{A_i\}_2$ можно получить много разбиений $\{A_i\}_2^h$ с числом классов $n_2^h = n_1$. Следовательно, можно говорить о мерах связи между одной ранжировкой P_1 и множеством ранжировок $\{P_i\}_2^h$.

В дальнейшем будем иметь в виду только строгие ранжировки с одинаковым числом элементов (в этом случае $n = N$), тогда $\mathcal{P}$ -множество всех ранжировок из n элементов $|\mathcal{P}| = n!$.

Мера связи определяется на множестве $\mathcal{P}^2$ и может либо приписывать паре $\langle P_i, P_j \rangle \in \mathcal{P}^2$ вещественное число, отражающее степень связи ранжировок, либо указывать только порядок на этих парах, т.е. говорить о том, слабее или сильнее связаны ранжировки друг с другом. Мы будем рассматривать только меры связи, приписывающие каждой паре ранжировок некоторое вещественное число.

Мы полагаем, что для разных целей требуются меры связи по-разному реагирующие на изменения одних и тех же характеристик ранжировок, поэтому предпочитаем построение параметрического множества мер связи, так чтобы свойства меры существенно зависели от параметров и, изменяя параметры, можно было бы получать меры с различными свойствами.

Таким образом, мера связи - это функция σ, определенная на $\mathcal{P}^2$, зависящая от двух переменных (ранжировок P_i и P_j) и некоторых параметров θ и принимающая значения в промежутке $[0, I]$. Вообще говоря, вопрос о нормировке меры не является принципиальным, важно только помнить, что экстремальные значения меры должны отвечать наиболььшим или наименьшим связным ранжировкам.

Можно рассматривать симметричные и несимметричные меры связи. Наиболее важным является вопрос о том, на каких парах ранжировок мера достигает максимального и минимального значений. Очевидно, от принятой гипотезы о совершенно связных и совершенно несвязных ранжировках должно зависеть как количество нулей и единиц у меры связи, так и определение пар объектов, на которых мера достигает экстремальных значений.

Перечислим возможные варианты.

I. Мера связи $\sigma(P_i, P_j)$ достигает I, когда

а) $P_i = P_j$ (только на тождественных ранжировках);

б) $P_i = P_i \vee P_j = \bar{P}_i$ (на тождественных и обратных ранжировках);

в) $P_j \in P_i^*$, где P_i^* - некоторый класс, ранжировки которого определены как совершенно связные с P_i.

П. Мера связи $\sigma(P_i\ P_j) = 0$ только когда

а) $P_j = \bar{P}_i$ - только на обратных ранжировках;

б) $\sigma(P_i, P_j) = \sigma(P_i, \bar{P}_j)$ или P_i, P_j такие, что для них дости -

гается $\min\limits_{\mathcal{P}^2} \{\sigma(P_i,P_j) - \sigma(P_i,P_j)\}$, т.е. значения меры связи ранжировки P_i с ранжировкой P_j и с обратной ей ранжировкой $\bar{P}_j$ равны или минимальны на множестве $\mathcal{P}^2$;

в) $P_j \in P_i^*$, где P_i^* – некоторый класс, ранжировки которого определены как совершенно несвязные с P_i.

Таким образом, возможны 8 различных формулировок требований к максимуму и минимуму меры.

Следующий важный вопрос: какими свойствами должны обладать меры связи между ранжировками? То есть, если заданы две ранжировки P_i и P_j с фиксированным значением меры связи между ними σ_{ij}, то необходимо перечислить возможные преобразования P_j при фиксированном P_i и указать, как должны изменяться при этом значения меры связи.

Как перечислить все возможные преобразования ранжировок, имеющие теоретическое и практическое значение – пока неясно. Но надо отдавать себе отчет, что это один из центральных вопросов построения теории мер связи.

Будем различать преобразования тривиальные при фиксированном числе элементов n и нетривиальные, связанные с изменением n. Ранжировки P_i и P_j без потери общности можно записать в алфавитном порядке $P_i = \langle 1,2,\dots,n \rangle$, $P_j = \langle 1,2,\dots,n \rangle$.

Поставим следующие вопросы: как должна изменяться мера связи между ранжировками P_i и P_j, если:

I) Проведена одна элементарная инверсия соседних элементов:

а) в начале ранжировки $P_j' = \langle 2,1,3,\dots,n \rangle$,

б) в середине $P_j'' = \langle 1,2,\dots,k{+}1,k,\dots,n \rangle$,

в) в конце $P_j''' = \langle 1,2,\dots,k,k{+}1,\dots,n,n{-}1 \rangle$.

2) Проведена циклическая перестановка одного элемента:

а) правая $P_j = \langle 1,2,\dots,n \rangle \to P_j' = \langle n,1,2,\dots,n{-}1 \rangle$,

б) левая $P_j = \langle 1,2,\dots,n \rangle \to P_j'' = \langle 2,3,\dots,n,1 \rangle$.

3) Проведено фиксированное количество элементарных инверсий, пусть их будет k :

а) один элемент в P_j имеет только одну инверсию, т.е. P_j имеет вид $P_j' = \langle 2,1,4,3,\dots,k{+}1,k,\dots,n,n{-}1 \rangle$;

б) один элемент имеет несколько инверсий $P_j'' = \langle 2,3,4,1,\dots$
$\dots, n \rangle$;

в) один элемент имеет k инверсий $P_j''' = \langle 2,3,\ldots,k-1,1,\ldots$ $\ldots, n \rangle$.

4) Зафиксируем некоторый элемент алфавита a_0 в P_i, при любой перестановке других элементов приписываем элементу a_0 в P_j номер k, где $k = 1,\ldots,n$.

Если необходимость различения случаев I) и 2) с помощью меры связи во многих случаях кажется очевидной, то случаи 3) и 4) требуют пояснения.

В 3-м случае нас интересует различение элементарных инверсий ("путаются" только наиболее близкие элементы) от любых остальных.

В 4-м случае нас интересует существование возможности слежения за некоторым особо выделенным элементом.

Рассмотрим нетривиальные преобразования.

I. Добавим элемент с номером $n+1$ в конец ранжировки P_i, так что $P_i' = \langle 1,2,\ldots,n,n+1 \rangle$ и

а) в начало ранжировки $P_j \to P_j' = \langle n+1,1,2,\ldots,n \rangle$,

б) в середину $P_j'' = \langle 1,2,\ldots,n+1,\ldots,n \rangle$,

в) в конец $P_j''' = \langle 1,2,\ldots,n,n+1 \rangle$.

Как должно зависеть изменение значения меры от числа элементов n?

2. Пусть имеются ранжировки P_{i1} и P_{i2}, P_{j1} и P_{j2} с числом элементов n_{i1} и n_{i2}, n_{j1} и n_{j2}, и меры связи $\sigma_{i1,i2}$, $\sigma_{j1,j2}$ соответственно.

Составим новые ранжировки P_i и P_j путем склеивания ранжировок P_{i1} и P_{i2}, и P_{j1} и P_{j2}, где I-й элемент ранжировок P_{i2} (P_{j2}) получает номера $n_{i1}+1$ ($n_{j1}+1$).

Как должны соотноситься значения мер связи σ_{i1i2}, σ_{j1j2} и σ_{ij}? Очевидно, перечисление всех возможных нетривиальных преобразований и возможного изменения при этом значений меры связи требует специального рассмотрения. В дальнейшем будем иметь в виду только тривиальные преобразования.

Рассмотрим с предложенных позиций наиболее распространенные конструкции, которые можно использовать как меры связи между ранжировками. Мера, введенная Кендэлом [I], представляет собой линейную функцию минимального числа перестановок в соседних элементов, необходимых для перехода от ранжировки P_j к P_i:

$$\tau_{ij} = 1 - \frac{2s}{1/2n(n-1)},$$

где n - число элементов. Если мера τ может быть определена как на номерном, так и на матричном описании, то мера Спирмена [I,4] определяется только на номерном описании. При расчете коэффициента ранговой корреляции ранжировок

$$\rho_{ij} = 1 - \frac{6\sum\limits_k (N_k^i - N_k^j)^2}{n(n^2-1)}$$

каждая инверсия между элементами оказывается взвешенной, вес инверсии увеличивается по мере отдаления элементов друг от друга. Обе меры принимают значения от -I до +I, $\tau = \rho = 1$ на тождественных ранжировках, $\tau = \rho = -1$ на обратных. Мера τ принимает значение, равное 0, только при четном значении величины $\frac{1}{2}n(n-1)$ - это будут ранжировки, различающиеся на $s = \frac{1}{4}n(n-1)$ инверсий; ρ принимает значение 0, когда величина $\frac{n(n^2-1)}{6}$ - четная. Таким образом, имеется в виду следующая гипотеза о связи ранжировок: максимально положительно связные - тождественные, максимально отрицательные связные - прямая и обратная, максимально несвязные - прямая и некоторая "средняя", мера сходства τ или ρ с которой равна или близка к нулю. Любое промежуточное значение оценивается статистическими методами на "существенность" при выбранном уровне значимости.

Кемени [2] предложил аксиоматическое введение меры Кендэла на множестве матриц смежности. Мера d удовлетворяет аксиомам, аналогичным аксиомам расстояния, является мерой близости ранжировок, принимает минимальное значение $d = 0$ на тождественных ранжировках и максимальное значение $d_{max} = n(n-1)$ на матрицах, соответствующих прямой и обратной ранжировкам. Таким образом, максимально близкие - тождественные ранжировки, максимально далекие - прямая и обратная. Любая одна элементарная инверсия вызывает одинаковое изменение значения мер. И следовательно, различные случаи тривиальных преобразований I-4 с помощью рассмотренных мер различать нельзя. Для того, чтобы иметь возможность взвешивать элементарные инверсии в зависимости от ранга или номера переставляемых объектов, вводятся структурные меры на упорядоченных разбиениях, описанных графами [3].

Задавая множество параметрических мер связи, исходим из следующих представлений. Пусть имеется некоторая эталонная ранжировка, верная по определению, и некоторая косвенная, порядок расположения элементов в прямой ранжировке принят за алфавитный $P_0 = \langle 1, 2, \ldots, n \rangle$. Любой элемент k в косвенной ранжировке P_i может получить тот же самый номер k, а также больший или меньший номер. Считаем, что каждая ошибка имеет свою цену. Можно составить матрицу цен ошибок C, где c_{kl} – цена приписывания элементу k номера l. Матрица цен ошибок является теоретической априорной конструкцией и зависит не от конкретного материала, а только от теоретических представлений о целях задачи и о природе элементов. Можно построить различные матрицы цен ошибок, исходя из различных теоретических предположений. Перечислим хотя бы некоторые разумные предположения и альтернативы, относительно значений C.

I) Очевидно, $c_{kk} = 0$.

2) $c_{kl} = c_{lk}$ или $c_{kl} \neq c_{lk}$, т.е. матрица цен ошибок симметрична, если цены ошибок завышения и занижения равны, и несимметрична, если эти цены различны.

3) Цена ошибки c_{kl} должна прежде всего зависеть от величины ошибки, т.е. от разницы номеров $|k - l|$ элемента в эталоне P_0 и в косвенной ранжировке P_f. Это может быть как угодно сложная зависимость: убывающая, возрастающая, колеблющаяся и так далее.

Будем рассматривать вариант, который, по-видимому, будет отвечать большинству содержательных задач, когда функция $f(|k - l|)$ монотонно-возрастающая и $f(0) = 0$. Если цена элементарной ошибки $f(k+1, k)$ зависит от общего числа элементов n, то f, очевидно, должна быть нормирована так, чтобы $c_{max} = 1$. Это может быть, например, семейство параметрических степенных функций $f = q\left(\frac{k-1}{n-1}\right)^{\alpha}$, $\alpha \geq 0$, при $\alpha = 0$ имеем вырожденный случай, когда любая ошибка имеет одинаковую цену. Для нормированной функции, очевидно, $0 \leq q \leq 1$.

4) Цена ошибки c_{kl} может также зависеть от k, тогда разумно потребовать, чтобы функция $\varphi(k)$ была монотонно-убывающей, с максимальным значением при $k = I$, например, $\varphi = k^{-\beta}$, $\beta \geq 0$. Таким образом, один из возможных вариантов матрицы C может быть записан следующим образом:

$$\underset{k>l}{c_{kl}} = q_1 \left(\frac{k-l}{n-1}\right)^{\alpha_1} \cdot \frac{1}{l}^{\beta_1} \text{ – ошибка завышения,}$$

$$c_{\substack{kl\\k<l}} = q_2\left(\frac{k-1}{n-1}\right)^{\alpha_2} \cdot \frac{1}{k}^{\beta_2} \text{ – ошибка занижения.}$$

Построим для косвенной ранжировки P_i матрицу Γ_i такую, что γ^i_{kl} равна частоте встречаемости варианта: k – номер элемента в эталоне, l – номер этого же элемента в P_i. Очевидно, для строгих ранжировок γ_{kl} = 0 или 0.

Определим суммарную ошибку завышения ранжировки P_i по отношению к P_0 как

$$C^+_{i0} = \sum_{k>l} \gamma^i_{kl} \cdot c_{kl}, \quad k \in P_0, \quad l \in P_i,$$

и суммарную ошибку занижений

$$C^-_{i0} = \sum_{k<l} \gamma^i_{kl} \cdot c_{kl}.$$

Введем показатель связи ранжировок P_0 и P_i $\varkappa_{i0} = C^+_{i0} + C^-_{i0}$, $\varkappa$ определен для всех строгих ранжировок; $\varkappa = 0$, когда $\gamma_{kl} = 0$ при $k \neq l$, так как все $c_{kk} = 0$ по построению, т.е. на тождественных ранжировках. В выбранном классе матриц C показатель связи $\varkappa$ достигает максимума, когда $\gamma_{kl} = 1$ при $k+l = n+1$ и $\gamma_{kl} = 0$ в других случаях, т.е. на обратных ранжировках. Показатель $\varkappa$ несимметричен по отношению к ранжировкам P_i P_j, т.е. $\varkappa_{i0} \neq \varkappa_{0i}$ по построению, при различных параметрах завышения и занижения, и симметричен при равных параметрах завышения и занижения.

Будем строить множество параметрических мер связи между ранжировками как некоторые функции σ от показателя, например,

1) $\sigma = 1 - \dfrac{\varkappa}{\max\limits_{\mathcal{P}} \varkappa}, \quad 0 \le \sigma \le 1;$

2) $\sigma = 1 - \dfrac{2\varkappa}{\max\limits_{\mathcal{P}} \varkappa}, \quad -1 \le \sigma \le 1.$

По желанию функцию σ можно отнормировать любым способом, но нам первый способ представляется наиболее верным методологически. Свойства меры σ вытекают из вышеразобранных свойств показателя $\varkappa$, но можно определить дополнительные симметричные меры связи σ^+ и для несимметричных $\varkappa$ $\sigma^+_{ij} = 1 - \dfrac{\varkappa_{ij} + \varkappa_{ji}}{2}$.

Цель проведения классификации мер связи состоит в получении таких таксонов (типов, классов и т.п.), внутри которых меры обладали бы достаточно близкими свойствами, а меры,принадлежащие разным таксонам,по некоторым существенным свойствам различались, причем чем выше уровень деления,тем более принципиальными свойствами должны различаться меры.

Классификация мер σ, или удобнее говорить о классификации показателей $\varkappa$, основана на классификации матриц цен ошибок C.

Предлагается выделять три типа показателей по соотношению параметров q_1 и q_2. I-й тип – $\varkappa$, построенные на симметричных матрицах: $q_1 = q_2$, $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$. Выделение классов внутри типов проведем по силе зависимости $\varphi(k)$, т.е. по значению параметра β. Выделение родов проведем по виду зависимости цены ошибки от величины ошибки, т.е. по значению параметра α (табл.I).

Таблица I

Выделение классов и родов в симметричном типе показателя $\varkappa$

Род	Класс по виду зависимости $\varphi(k)$		
	$\beta=0$	$0<\beta\leq1$	$\beta>1$
α = I	I	2	3
$\alpha = 2$	4	5	6
$\alpha > 2$	7	8	9

Ко 2-му типу отнесем $\varkappa$,построенные на несимметричных матрицах, в которых параметр завышения q_1 имеет больший вес по сравнению с параметром занижения q_2, к третьему типу отнесем $\varkappa$, для которых $q_1 < q_2$. В каждом типе выделим подтипы по возможным соотношениям параметров α_1 и α_2, β_1 и β_2 (табл.2).

Выделение классов и родов в I-м подтипе проводится аналогично I-му типу, в остальных подтипах по меньшим параметрам, считая больший параметр всегда много больше 2.

Таблица 2

Выделение подтипов для несимметричных типов показателя $\varkappa$ для второго ($q_1>q_2$) и третьего ($q_1<q_2$) типов

	$\alpha_1=\alpha_2$	$\alpha_1<\alpha_2$	$\alpha_1>\alpha_2$
$\beta_1=\beta_2$	I	2	3
$\beta_1<\beta_2$	4	5	6
$\beta_1>\beta_2$	7	8	9

Не будем сейчас останавливаться на полной классификации перечислений по всем возможным соотношениям параметров в типах 2 и 3, укажем только, что получены новые виды мер связи,в частности,меры связи, обладающие нужными свойствами для различения тривиальных преобразований I-4.

Общая формула $\varkappa$ для I-го типа

$$\varkappa_{ij} = \sum_{k>1}\left(\frac{k-1}{n-1}\right)^{\alpha}\cdot\frac{1^{\beta}}{1} + \sum_{k<1}\left(\frac{1-k}{n-1}\right)^{\alpha}\cdot\frac{1^{\beta}}{k} .$$

Заметим сразу, что показатели $\varkappa$ I-го вида являются основой построения стандартного коэффициента Спирмена

$$\rho'_{ij} = 1 - \frac{\sum_k |N_k^i - N_k^j|}{E^{n^2/2}} ,$$

а показатели $\varkappa$ 4-го вида для построения коэффициентов ранговой корреляции Спирмена $\rho(2)$.

По построению $\varkappa$ параметр β ответственен за различение инверсий соседних элементов в зависимости от того, какие именно элементы переставляются. Например, если нужно отличить инверсию I-го элемента от всех остальных, наилучшего отделения можно добиться при $\varkappa$ 3-го вида, где $\beta \gg 1$. В этом случае из $n!$ ранжировок выделяется класс, включающий n ранжировок, в которых I-й объект имеет номер I.

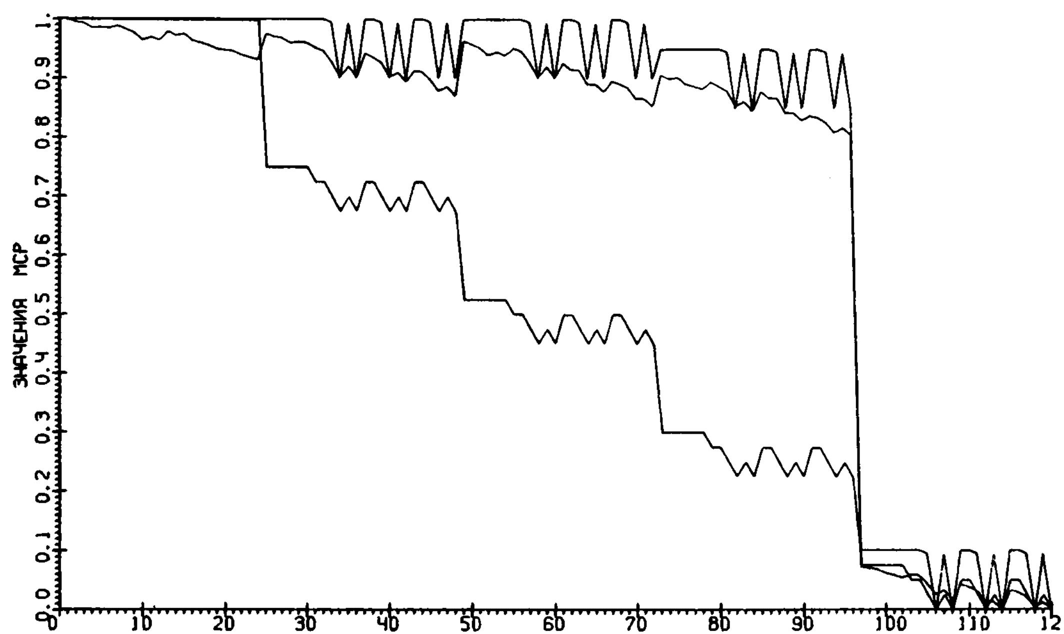
Элементарные нарушения (3-е тривиальное преобразование) отличаются от любых других независимо от числа элементов n в ранжировке при показателе $\varkappa$ I-го типа 3-го вида при $\alpha \log_2 n$ и ненормированной цене ошибки c_{k1}.

Правая и левая циклические перестановки различаются при любом виде показателей 2-го и 3-го типа. Какое именно значение параметров и их соотношение выбрать, зависит от того, какой ошибке придается больший вес.

На графике приведены значения меры связи при $n = 5$ для $n!$ ранжировок для некоторых видов σ 3-го типа.

Можно доказать, что:

I) Меры связи I-го подтипа 3-го вида при $\beta \gg 1$ позволяют разбить множество из $n!$ ранжировок на n классов, различающихся на $\Delta \approx \frac{1}{n-1}$, где каждый k-й класс будет содержать только ранжировки, в которых I-й элемент алфавита имеет номер k (или совершенно симметрично мерой связи 2-го типа выделять классы ранжировок, где каждый k-й класс содержит ранжировки, в которых k-й элемент алфавита получил номер I).



Мера связи σ 3-го типа $q_1 < q_2$, $q_1 = 0.I$, $q_2 = 0.9$ для n' ранжировок, при $n = 5$.
График I – σ 2-го подтипа, 2-го вида, $\alpha_1 = I$, $\alpha_2 = I0$, $\beta_1 = \beta_2 = 1$.
График 2 – σ I-го подтипа, 3-го вида $\alpha_1 = \alpha_2 = 1$, $\beta_1 = \beta_2 = 10$.
График 3 – σ I-го подтипа, 6-го вида $\alpha_1 = \alpha_2 = 10$, $\beta_1 = \beta_2 = 10$, пояснения на стр. 76,78.

2) Меры связи I-го подтипа 6-го вида, где α имеет достаточно высокую степень, позволяют разбить все множество ранжировок на два класса, где в один класс входят ранжировки, в которых I-му элементу алфавита присвоен номер n, а в другой – все остальные ранжировки (или, симметрично, мерой связи 2-го типа выделять два класса ранжировок, I-й с ранжировками, в которых последнему элементу алфавита присвоен номер I, во 2-й – все остальные). Очевидно, этот вид мер даст и максимальное различие для правой и левой циклических перестановок.

Таким образом, с помощью новых мер связи появляются новые возможности для формулировки гопотез о связности ранжировок.

В заключение скажем несколько слов о постановке задач на выбор наиболее подходящей меры для конкретной задачи.

Задача выбора меры связи – это, как правило, вспомогательная задача, встроенная в некоторую "главную" задачу анализа данных и следовательно, имеющая некоторый внешний критерий эффективности.

Из анализа содержательной задачи формулируются априорные предположения о виде связи между ранжировками, которые должны позволить выбрать тип, подтип, класс и род меры связи. Очевидно, это неформальный этап выбора меры. Если никаких априорных предположений нет, то должна выбираться наиболее простая зависимость.

Экспериментальный материал может быть задан либо в виде частичного квазипорядка на множестве ранжировок, либо в виде значений меры для некоторых пар ранжировок $\lambda(P_i, P_j) = \lambda_{ij}$, $i = \overline{1, I}$, $j = \overline{1, J}$, тогда требуется подобрать σ_k такую, что $\sigma_k(P_i, P_j) - \lambda_{ij} \leq \epsilon$, $\forall i, j$. Пока трудно говорить о постановке задачи выбора меры, согласованной с любым заданным квазипорядком или любыми заданными значениями λ. Можно говорить только об интерпретации заданного квазипорядка в терминах матрицы цен ошибок и определении для него соответствующего типа, класса, рода и т.д. показателя μ. Определение численных значений параметров q, α, β может быть сделано обычными методами приближения функции к эмпирическим данным. В общем случае система уравнений для нахождения параметров может быть достаточно сложной.

Поиски эффективной меры, т.е. меры, отвечающей некоторому внешнему критерию, также сводятся к решению задач построения меры по теоретическим предположениям и эмпирическим данным. Потому что любой критерий, зависящий от меры связи, очевидно, зависит либо от квазипорядка на ранжировках, порожденного мерой, либо от численных значений выбранной меры.

Литература

1. КЕНДЭЛ М. Ранговые корреляции. -М.: Статистика,1975.-214с.

2. КЕМЕНИ Дж., СНЕЛЛ Дж. Кибернетическое моделирование.Некоторые приложения. -М.: Сов.радио, 1972. - 192 с.

3. ЛИТВАК Б.Г. Экспертная информация.Методы получения и анализа. -М.: Радио и связь, 1982. - 183 с.

4. Экспертные методы оценки качества промышленной продукции. ГОСТ 23554.2-81, М., 1982, 66 с.

Поступила в ред.-изд.отд.
9 апреля 1984 года