

## **Система анализа нечисловой информации «САНИ»**

**С.Ю.Адамов**

**(Москва)**

Описывается система, предназначенная для анализа нечисловых и смешанных данных. Приводятся примеры применения « реальной эмпирической информации.

Ключевые слова: нечисловая информация, пакет программ, компьютерный анализ, анализ соответствий , матриц» сопряженности, модель квазисимметричности.

Новые типы вычислительной техники, и прежде всего персональные компьютеры (ПК), позволяют проводить оперативный и более тонкий анализ информации. Теперь диалог специалиста в конкретной области с вычислительной машиной может проходить без посредничества программиста и операторов. Новые графические средства обеспечивают наглядное представление результатов применения различных

методов, что значительно облегчает их восприятие. Однако для реализации преимуществ ПК необходимо специальное программное обеспечение. Практика показывает, что адаптация пакетов программ, созданных для больших ЭВМ, на ПК (SPSS, BMDP, SPAD) не обладает должным уровнем «дружелюбия». Использование таких продуктов требует долгого обучения и солидной подготовки в области математической статистики. При этом следует отметить, что несмотря на широту имеющихся методов и программ, относительно слабо продвинуто предварительный анализ данных. Между тем, специфика социологических данных (неоднородность, зависимость наблюдений друг от друга и от времени, обилие качественных признаков) в первую очередь требует его применения. Имеются в виду методы, не опирающиеся на вероятностные предположения о природе данных, например детерминационный анализ и анализ соответствий<sup>1</sup>, которые в отличие от классических методов математической статистики дают результаты, относящиеся к выборке, а не к генеральной совокупности. Их можно эффективно использовать как на предварительном этапе изучения данных (для выявления аномальных наблюдений и возможных кластеров), так и для интерпретации результатов моделирования.

Кроме того, отсутствует адекватное статистическое и программное обеспечение анализа динамики качественных показателей, который особенно актуален в связи с проводимыми в последнее время лонгитюдными обследованиями.

Эти соображения легли в основу развития некоторых методов и создания Системы анализа нечисловой информации (САНИ).

---

<sup>1</sup> Последний, будучи чрезвычайно популярен за рубежом, не имеет адекватного отражения в отечественной литературе и практике исследований.

## 1. Общие сведения

Система анализа нечисловой информации<sup>2</sup> предназначена для обработки данных, измеренных в разнородных шкалах: номинальной, порядковой, интервальной и количественной. В первую очередь она ориентирована на социально-экономические задачи, но может использоваться в биологии, медицине, других областях, где приходится иметь дело с нечисловой информацией.

САНИ реализована на совместимых с IBM персональных компьютерах и требует около 1 МБайта памяти на жестком диске, MS DOS не ниже версии 3.2 и адаптер типа Hercules, CGA, EGA или VGA. Она предполагает лишь элементарное знакомство пользователя с ПК. Работа осуществляется в диалоговом режиме с помощью иерархического меню или непосредственно нажатием определенных комбинаций клавиш.

Система позволяет одновременно обрабатывать до 320 признаков (вопросов); число объектов (респондентов) не должно превосходить 32 000 для категоризованного признака и 8 000 - для числового. Однако приведенные ограничения, имеют чисто технический характер, и допустимый объем данных зависит от используемого метода и типа компьютера.

В любой момент доступны: справка об используемом методе со ссылкой на литературу или разъяснение сложившейся ситуации; справки обо всех активных переменных, содержащие информацию, полученную от пользователя при

---

<sup>2</sup> В разработке системы помимо автора принимали участие А.Г.Шимельмиц, А.А.Халиев, М.М.Левинсон и А.А.Кошелев. Авторы опирались на идейную и организационную поддержку С.А.Айвазяна.

первом вводе, и некоторые результаты проделанного анализа; данные об объектах (соответствующие значения переменных). Результаты анализа выводятся на экран и могут быть распечатаны или скопированы в файл для последующего включения в отчет.

## **2. Функциональное наполнение**

Методы, используемые САНИ, распадаются на три группы. Первая — реализует возможности базы данных, вторая - объединяет средства предварительного анализа, позволяющие сформировать гипотезы о структуре данных, выявить «выбросы». При этом эффективно применяются графические возможности современной вычислительной техники. Методы, входящие в третью группу, используют вероятностные предположения о природе данных и позволяют проверять различные гипотезы. Особое внимание уделено анализу нечисловых признаков, изменяющихся во времени.

*Работа с данными.* Они могут быть введены вручную, импортированы из прямоугольных таблиц «объект-признак» или «признак-объект» в кодах ASCII или из общего статистического пакета SYSTAT. Имеется возможность экспортировать данные в виде таблиц ASCII или в системном формате SYSTAT.

В системе «САНИ» каждому признаку соответствует справка, содержащая сведения о шкале, в которой он измерен, код и число пропущенных значений, имена и частоты категорий, историю создания и комментарии пользователя. Кроме того, в справке хранятся некоторые результаты предыдущей обработки: имена независимых и тесно связанных с данной переменных. Они предостерегают исследователя от мало осмысленных шагов, например от использования независимых переменных в анализе соответствий.

Имеется возможность получать подвыборки: отбирать или удалять объекты с фиксированной комбинацией значений переменных. Можно создавать новые переменные в виде фиксированных комбинаций значений имеющихся переменных, агрегированием категорий, разбиением количественных переменных на интервалы, всевозможными комбинациями двух переменных (для снижения размерности). Все преобразования фиксируются в справках. Мощный редактор позволяет вводить и проверять данные вручную, а также изменять имя, комментарий, шкалу и значения переменных.

Допустимы два режима функционирования системы: при небольшом объеме данных вся интересующая исследователя информация помещается в оперативной памяти ПК; в противном случае значения признаков читаются с твердого или гибкого диска по мере необходимости.

**Предварительный анализ.** Процедура построения гистограмм позволяет в удобном виде изображать и сравнивать распределения признаков. Для признаков, измеренных в номинальной шкале, имеется возможность переставлять категории согласно возрастанию или убыванию частот или так, чтобы приблизиться к нормальному распределению; проверять гипотезу о нормальности, а также создавать числовые переменные, значениями которых будут квантили на интервалах стандартизованной нормально распределенной случайной величины, соответствующих категориям исходного нечислового признака.

При детерминационном анализе на экран выводится дерево распределения объектов по категориям качественных переменных. Переменные и категории могут выбираться вручную или автоматически по критерию максимизации мер связи.

Имеется возможность последовательного определения переменных и их значений для наилучшего предсказания.

Пусть требуется найти комбинацию значений признаков  $x_1, \dots, x_p$  для

наилучшего предсказания категории  $k$  признака  $y$ . Выбор «предсказателей» производится последовательно. Пусть значения выбранных  $s$  признаков фиксированы. Следующий признак  $x_t$  и соответствующая категория  $j$  определяются из условия максимума по  $t$  и  $j$  выражения

$$\lambda_{tj} = \frac{f_{jk}^{(t)}}{f_{j.}^{(t)}} - \frac{f_{.k}^{(t)}}{f_{..}^{(t)}}$$

где  $f_{jk}^{(t)}$  элемент слоя  $s+2$ -мерной таблицы сопряженности, в котором пересекаются значения признаков  $x_t$  и  $y$  при фиксированных  $x_{i1} \dots x_{is}$ ;  $f_{j.}^{(t)}, f_{.k}^{(t)}, f_{..}^{(t)}$ , сумма (по индексу,

обозначенному точкой). Величина  $\lambda_{it}$  показывает, насколько знание  $j$ -и категории  $x_t$  улучшает предсказание  $k$ -й категории  $y$ , если  $x_{i1}, \dots, x_{is}$  известны.

Допустимо представление изменения состояний объектов в виде ломаных линий. Пользователь может получить информацию о числе и номерах объектов в каждом состоянии, выделить их траектории, проходящие через фиксированные состояния.

Процедура выявления эталонных объектов позволяет найти наиболее распространенные комбинации значений выбранных признаков, обнаружить многомерные группировки.

Обеспечивается анализ двумерных таблиц сопряженности. Вычисляются и выводятся на экран частоты, проценты, отклонения от независимости, средние значения количественной переменной, средние разности двух количественных переменных в каждой клетке и т.п. Можно представлять две таблицы одновременно, изменять формат.

Большое число объектов в некоторой клетке таблицы сопряженности может быть следствием высоких маргинальных частот, а не связи между соответствующими категориями. Ф.Мостеллером была предложена процедура уравнивания [1], позволяющая нивелировать этот эффект. Суммы по строкам и столбцам одновременно приводятся к стандартной величине, после чего по элементам преобразованной таблицы можно судить о близости категорий.

Определяются некоторые стандартные меры парной связи с учетом шкал, в которых измерены признаки и, где это возможно, ошибки I рода в предположениях независимости и полной связи. Результаты фиксируются в справках. Допустимо построение матриц, содержащих коэффициенты парной связи нечисловых признаков.

Реализованы оригинальные алгоритмы выделения блочной и квазиблочной структур. В первом случае строки и столбцы таблицы сопряженности переставляются так, чтобы ненулевые элементы образовали непересекающиеся блоки, расположенные вдоль главной диагонали [2]. Во втором случае перестановка строк и столбцов производится с целью минимизации величины

$$\sum_{i,j} f_{ij}(i-j)^2,$$

где  $f_{ji}$  - элемент анализируемой таблицы. Как правило, получаемые блоки имеют непустые пересечения. Если таблица сопряженности недостаточно «разреженная», то, задав пороговое число наблюдений, элементы с меньшим числом наблюдений можно приравнять нулю. Очевидно, что блочная структура эквивалентна двумерному разбиению элементов на непересекающиеся классы. При квазиблочной структуре один элемент может принадлежать двум соседним классам.

Многомерные таблицы сопряженности выводятся на экран по слоям.

Анализ соответствий [3] позволяет получать в евклидовом пространстве одновременное представление объектов и категорий качественных переменных. Ищется такая проекция данных в пространство небольшой размерности, чтобы приблизить следующую структуру: каждая точка, соответствующая объекту, расположена в центре облака характерных для него категорий, а точка, представляющая категорию переменных, в свою очередь, являлась центром тяжести соответствующих объектов. Математически задача сводится к определению

собственных векторов некоторой матрицы. Качество представления оценивается долей следа, объясняемой факторами (это заниженная оценка, особенно когда число переменных больше двух).

**Моделирование.** На стадии моделирования предоставляются возможности проверки гипотез о независимости, квазинезависимости (в таблице имеются клетки, в которые по содержательным соображениям не могут попасть наблюдения), симметричности, маргинальной однородности и квазисимметричности [4] таблиц сопряженности. Последняя модель записывается следующим образом

$$f_{ij} = a_i b_j s_{ij}, \quad s_{ij} = s_{ji}, \quad \sum_i a_i = \sum_j b_j = 1, \quad (1)$$

где  $f_{ij}$  — элемент рассматриваемой таблицы;  $a_i, b_i, s_{ij}$  — параметры.

Для дискретных случайных процессов проверяются гипотезы однородности (независимости переходной матрицы от времени) и марковости. Последняя, как правило, требует оценки очень большого числа параметров, поэтому в САНИ она сведена к проверке условной независимости в моменты  $t-1$  и  $t+1$  при всех фиксированных состояниях в момент  $t$ . Эта процедура может применяться и для определения графа связи [5].

При проверке любых гипотез и моделей определяется статистика согласия, уровень значимости, ожидаемая и наблюдаемые матрицы, а также вклады клеток в критерий согласия.

Имеется возможность рассчитывать такие характеристики марковских цепей, как вероятности состояний в заданные моменты времени, среднее время ожидания для однократного попадания, выявлять изолированные и поглощающие состояния.

Если процесс регулярен, определяется стационарное распределение, средние времена ожидания и первого попадания, предельные ковариации и корреляции состояний, показатели мобильности и связи между состояниями. Если имеются поглощающие состояния, то переходная матрица приводится к каноническому виду и вычисляются средние времена и вероятности поглощения [6].

Показатели, изменяющиеся во времени, можно разделять на заданное число классов согласно критерию активности (числу перемен состояния), а если состояния упорядочены - уровню состояний или направлению изменений (повышающиеся, неподвижные, понижающиеся, колеблющиеся). Реализован метод fe-средних, в котором расстояние между объектами определяется как сумма весов совпадающих категорий. Вес категории задается величиной, обратной ее частоте. Центры исходных классов образуют объекты с наиболее часто встречающимися комбинациями значений переменных. Любая из перечисленных выше классификаций или экзогенная качественная переменная могут служить началом итерационного процесса. Для полученного разбиения функция правдоподобия в предположении независимости от времени достигает, в общем случае, локального максимума внутри каждого класса.

Модель квазисимметричности оказывается наиболее продуктивной, когда имеется возможность сравнивать значения параметров для различных упорядоченных групп или в различные моменты времени.

Пусть  $f_{ijt}$  - ожидаемое согласно модели, число объектов, перешедших из состояния  $i$  в  $j$  в момент  $t$ . Предлагаются две модификации модели квазисимметричности (1), отличающиеся от классического случая [7] как формулировкой, так и методом оценки параметров [8]

$$f_{ijt} = a_{it} b_{jt} s_{ij}, \quad s_{ij} = s_{ji}, \quad \sum_{i,j} a_{it} = \sum_{j,t} b_{jt} = 1, \quad i \neq j$$

и

$$f_{ijt} = a_i b_j s_{ijt}, \quad s_{ijt} = s_{jit}, \quad \sum_i a_i = \sum_j b_j = 1, \quad i \neq j.$$

Диагональные элементы слоев матрицы обмена  $f_{ijt}$  не оцениваются. На содержательном уровне это означает, что в каждый момент времени  $t$  мы концентрируем внимание на объектах, переходящих из одного состояния в другое. С увеличением параметров  $a_{it}$  ( $a_i$ ) растет число объектов, покидающих состояние  $i$  в момент  $t$ . Параметры  $b_{jt}$  ( $b_j$ ) характеризуют приток в  $j$ . Эти величины называются показателями отталкивания и привлекательности соответственно. Матрица с элементами  $s_{ij}$  симметрична и задает индекс близости состояний (для второй модели индекс близости зависит от времени).

Проверяются предположения о равенстве распределений двух или нескольких количественных переменных с помощью ряда непараметрических критериев. Возможна проверка гипотез о постоянном сдвиге и гипотез для парных Наблюдений.

Имеется программа логлинейного анализа, которая позволяет проверять согласие моделей, включающих взаимодействия до десятого порядка. Выбор модели производится в удобном интерактивном режиме.

### 3. Экспериментальные расчеты

Для расчетов использовались данные о социальной мобильности 859 супружеских пар, проживающих в г. Таганроге [9]. Для каждой семьи регистрировалось число детей, образование супругов, социальные категории, к которым они принадлежали в первые десять лет брака. Применялась следующая типология: С1 - мужчины, находящиеся на службе в Советской Армии, женщины, временно не работающие или находящиеся в отпуске по уходу за ребенком; С2 - младший обслуживающий персонал; С3 - рабочие низкой квалификации; С4 - рабочие средней квалификации; С5 - рабочие высокой квалификации; С6 - специалисты со средним специальным образованием; С7 - планово-экономические и счетные

работники; С8 - специалисты с высшим образованием; С9 - руководители предприятий и отдельных подразделений» освобожденные партийные работники

**Изменение состояния семей во времени.** Состояние семьи характеризуется комбинацией состояний супругов. Соответствующая таблица сопряженности на первом и десятом годах брака имеет размеры 68 x 58 (пустые строки и столбцы удаляются в САНИ автоматически). Естественно, что в этой таблице многие клетки пусты или содержат по одному-двум респондентам. Процедура выявления блочной структуры при пороговом значении 4 порождает матрицу

	C4C2	C4C3	C4C4	C8C2	C5C3	C5C4	C5C	C8C	C8C
C3C3	5	5							
C4C2				8					
C4C3	6		6	4	11	6			
C4C4			12			7			
C4C6							7		
C5C6							5		
C8C6								6	
C8C8									5

Эта таблица показывает лишь наиболее типичные переходы. Клетки, не представленные в ней, содержат менее четырех объектов.

Первый блок объединяет семьи, в которых на первом году брака муж - рабочий низкой или средней квалификации (С3 или С4), а социальная категория жены не выше категории мужа (С2, С3, С4). К десятому году статус мужа, как правило, повышается на одну ступень, а жена либо не меняет категорию, либо переходит в сферу обслуживания.

Второй блок образован семьями, где муж - рабочий средней или высокой квалификации (С4 или С5), а жена работает на должности, требующей среднего специального образования (С6). В таких семьях к десятому году брака квалификация мужа повышается, а категория жены остается без изменений.

Еще имеются два немногочисленные блока семей, в которых муж - специалист с высшим образованием. В одном из них жена работает на должности, требующей среднего специального образования, а в другом - специалист с высшим образованием. Оба типа семей не изменяют свой статус к десятому году брака.

**Связь образовательных уровней супругов.** Соответствующая таблица сопряженности имеет вид

	Неполное среднее	Среднее	Среднее специальное	Высшее	Сумма
Неполное среднее	72	24	19	5	120
Среднее	22	24	15	3	64
Среднее специально	13	23	29	12	77
Высшее	5	12	35	46	98
Сумма	112	83	98	66	359

Строки определяют уровень образования мужа, а столбцы - образование жены. После уравнивания этой таблицы получилась следующая

	Неполное среднее	Среднее	Среднее специальное	Высшее	Сумма
Неполное среднее	54,04	22,12	16,41	7,45	99,99
Среднее	29,46	39,46	23,11	7,98	99,99
Среднее специально	13,21	28,70	33,90	24,22	100,0
Высшее	3,30	9,74	26,61	60,36	100,0
Сумма	100,0	100,0	100,0	100,0	

Значение коэффициента Крамера [10] равно 0,369, и гипотеза о независимости отвергается с вероятностью ошибки 0,002. Меры Гуттмана [11] приблизительно равны:  $\lambda_\alpha = 0,2385$ , а  $\lambda_\beta = 0,2389$ , поэтому нельзя говорить, что образование одного из супругов определяет образовательный уровень другого. Тем не менее, гипотеза о симметричности отвергается: статистика согласия равна 19,55 при 6 степенях свободы; ошибка I рода равна 0,0033. Гипотеза об однородности распределений мужа и жены также отвергается: статистика равна 16,01 при 3 степенях свободы, ошибка I рода равна 0,0011.

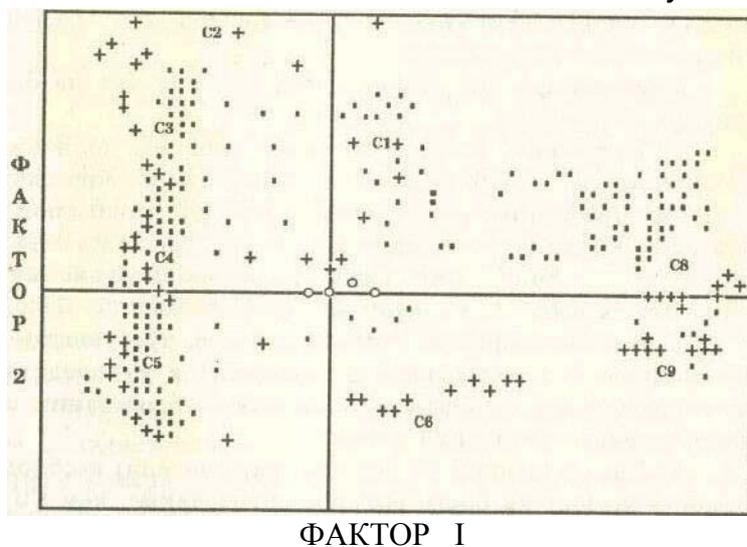
Опираясь лишь на исходную таблицу сопряженности, можно только догадываться о структуре взаимосвязей между категориями. В самом деле, большое значение (72) в клетке, связывающей супругов с низким образованием, может быть следствием многочисленности этой категории (120 среди мужчин и 112 среди женщин). Обычно в таких случаях рассчитывают проценты по строкам и по столбцам. Процедура уравнивания [1] нивелирует эффекты обеих маргинальных частот одновременно.

В преобразованной таблице максимумы по строкам и по столбцам расположены на главной диагонали, что говорит о преимущественной однородности семей. Однако наиболее тесная связь у категорий не с низким образованием, как можно было бы предположить согласно исходной таблице, а с высшим образованием. Числа, лежащие непосредственно под диагональю, превосходят симметричные им относительно диагонали. Это говорит о том, что если образовательный уровень супругов различается на одну ступень, то он выше, как правило, у мужа. Заметим, что чисел, для которых абсолютная разность номера строки и столбца больше единицы, кроме одного элемента, в нижней части матрицы меньше, чем в верхней. Отсюда можно сделать вывод: если разрыв между образованием мужа и жены больше одной ступени, то жена в

среднем имеет более высокий уровень. Исключение составляет случай, когда у мужа высшее, а у жены - среднее образование.

**Социальная мобильность мужчин.** Применение анализа соответствий к десяти переменным, отражающим такую мобильность, выявило «выброс»: только один объект находился в категории плановых работников (С7) и не менял ее в течение десяти лет. После его удаления первые два фактора стали объяснять 22% следа. Это довольно хорошая «подгонка» для множественного анализа соответствий. Ниже приводится проекция данных на эти факторы (см. рис.1).

Рисунок 1



Социальная мобильность мужчин: проекция данных на 1-й и 2-й факторы. Здесь: + - категории; • - объекты; o - проекции категорий дополнительной переменной (образования), не участвовавшей в определении факторов.

Все категории, кроме С1 (армия) и С2 (сфера обслуживания) образовали довольно плотные группы. Это говорит о том, что респонденты в среднем не склонны к смене состояний. В группах С6-С9 (специалисты со средним специальным и высшим образованием) соответствующие категории за небольшим исключением упорядочены по времени. Категории С2 и особенно С1 сильно разбросаны, следовательно, они являются переходными. Крестики, соответствующие категории С1 (армия) в различные моменты времени расположены; вблизи центра, что делает их одинаково доступными из других состояний (расстояние до С8 (инженеры) и С9 (руководители) несколько больше). Категории образования упорядочились слева направо (от неполного среднего до высшего), что позволяет интерпретировать первый фактор как объясняющий различия в образовании

Приведенная диаграмма рассеяния наводит на следующие предположения.

1) Категории, соответствующие рабочим должностям (С3—С5) требуют приблизительно одинакового образования так как их проекции на ось абсцисс примерно совпадают. Исключение составляет категория рабочих высокой квалификации на первом году, располагающаяся значительно правее группы таких рабочих в прочие моменты времени.

2) Для категорий С6-С9 (должности, требующие среднего специального образования, инженеры и руководители судя по их упорядоченности, требования к образованию ослабевают по мере увеличения стажа.

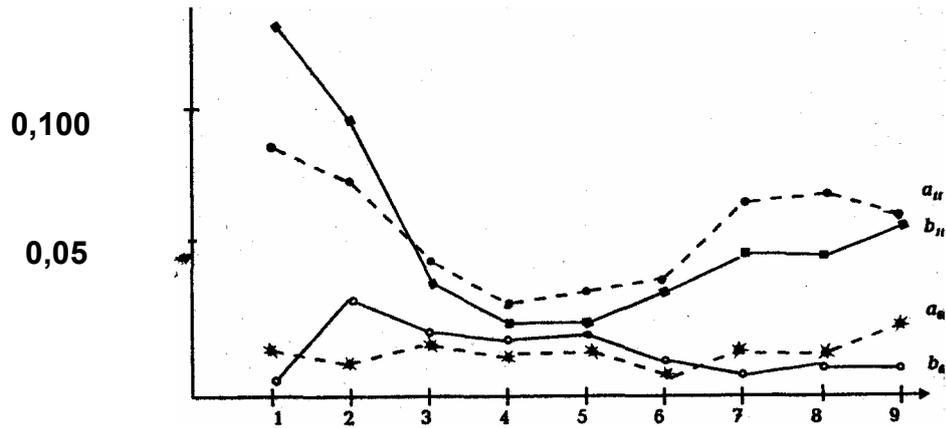
3) Для категории С8 (инженеры, учителя) необходим в среднем несколько более высокое образование, чем С9 (руководители): соответствующие точки расположены несколько правее. По-видимому, это объясняется неоднородностью категории С9. Результаты более подробного анализа подтверждают это: категории С6 и С8, строго говоря, не «требуют» соответственно среднего специального и высшего образования.

**Социальная мобильность женщин.** Рассмотрим результаты оценки параметров модели квазисимметричности. Статистика согласия равна 284,65 при 349 степенях свободы; уровень значимости высок: 0,995, хотя объем выборки слишком мал, чтобы можно было полагаться на распределение  $\chi^2$ .

Следуя интерпретации параметров миграционных моделей [12], естественно предположить, что показатели привлекательности  $b_{jt}$  и отталкивания  $a_{jt}$  выражают различные стороны индивидуальной функции полезности в зависимости от состояния и времени: уровень заработной платы, тяжесть работы, престиж. Индекс близости состояний характеризует легкость перехода из одного в другое (различия в профессиональной подготовке, в образовании). В данном примере он предполагается независимым от времени.

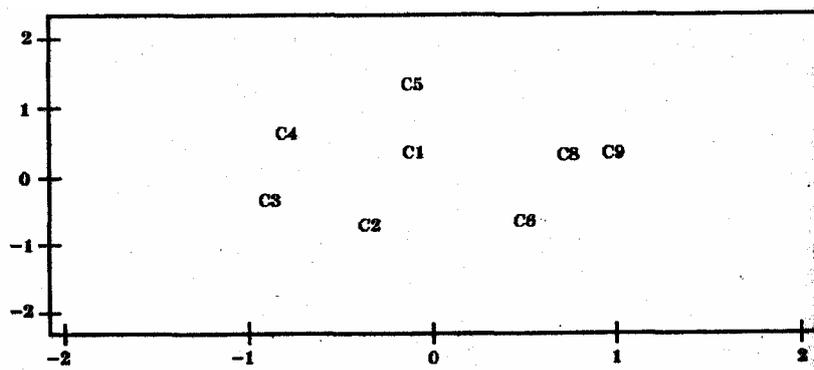
Для всех категорий наибольшие значения показателей привлекательности и отталкивания падают на первые годы брака. Будучи очень большими вначале, для С1 они уменьшаются к 3, 4 годам брака (см. рис.2). Следующий пик, связанный, по-видимому, с рождением второго ребенка, наблюдается в конце периода. Уровень динамики остальных показателей приблизительно на порядок меньше, чем  $a_{1t}$  и  $b_{1t}$ . Еще, пожалуй, можно выделить состояние С6, через которое в первые три года проходит значительное число респондентов. Остальные параметры изменяются аналогично показателям привлекательности и отталкивания для С6, но на более низком уровне. Можно заметить, что практически для всех состояний значения обоих показателей, начиная с пятого года, очень близки. Это говорит о том, что наступает равновесный обмен, т.е. для каждого состояния число выбывших приблизительно равняется числу прибывших (см. рис.2).

Рисунок 2



Социальная мобильность женщин: сравнительная динамика показателей отталкивания и привлекательности ; социальных категорий С1 и С6; пунктирные линии - динамика показателей отталкивания  $a_1$  и  $a_6$  для категорий С1 и С6 соответственно; сплошные линии - динамика показателе привлекательности  $b_1$  и  $b_2$  для С1 и С6 соответственно.

Рисунок 3



Социальная мобильность женщин: взаимная конфигурация восьми социальных категорий на плоскости, полученная методом многомерного шкалирования.

Ниже приводится результат применения многомерного шкалирования (алгоритм Крускала [18]) к матрице близости. Начальное значение стресса - 0,188 - в результате 20 итераций было уменьшено до 0,125. Состояния образовали круг, в центре которого оказалась категория, С1 (временно не работающие или находящиеся в отпуске по уходу за детьми), одинаково доступная из всех остальных категорий (см. рис.3).

Просматриваются две возможные карьеры: начиная с С2, через С3, С4 и С5 - рабочая и С6, С8, С9 - должности, требующие образования.

Исследования с помощью марковских моделей подтверждают сделанные выводы.

**Пример работы, метода классификации, лонгитюдных данных.**  
 Генерировались две однородные **марковские** цепи с общей переходной матрицей

0,5	0	0,5	0	0	0
0	1	0	0	0	0
0,5	0	0,5	0	0	0
0	0	0	0	1	0
0	0	0	0,5	0	0
0	0	0	0	1	0,5

Посредством агрегирования состояний  $i$  и  $i+3$  ( $I=1,2,3$ ) получалась смешанная цепь. Например, первое состояние смеси в каждый момент времени характеризовало объекты, находящиеся в первом состоянии 1-й или во втором состоянии 2-й цепи. Задача состояла в разделении этих объектов, В результате применения процедуры классификации (начальное разбиение содержало 30% неправильно классифицированных объектов) за три итерации точно были восстановлены исходные цепи.

## Литература

1. *Bishop Y., Fienberg S: Holland W.* Discrete Multivariate Analysis; Theory and Practice. Cambridge, 1976.
2. *Pages Caillez J. L/Introduction и l'analyse de donnees.* Paris, 1979.
3. *Lebart L., Morineau A., Warwick K.* Multivariate Descriptive Statistical Analysis. N.Y., 1984.
4. *Caussinus H.* Quelques points de vue sur l'analyse des tableaux d'echanges //Annales de L'INSEE. 1976. №№ 22,23.
5. *Гаврилец ЮЛ.* Социально-экономическое планирование. Системы и модели. М.: Экономика, 1974.
6. *Кемени Д., Снелл Д.* Конечные цепи Маркова. М.: Наука, 1970.
7. *Caussinus H., Thelot C.* Note complementaire sur l'analyse statistique des migrations //Annales de L'INSEE. 1976. №№ 22,23.
8. *Адамов С.Ю. Герасимова ИА.* Модели эволюции социального статуса супругов и ее факторы //Сб. научн. трудов ИЭ ОПП СО АН СССР. Новосибирск: ИЭ ОПП, 1990.
9. Народное благосостояние. Методология и методика исследования. М.: Наука, 1988.
10. *Кендалл МДж., Стьюарт А.* Статистические выводы и связи. М.: Наука, 1973.
11. *Антон Г.* Анализ таблиц сопряженности. М.: Финансы и статистика, 1982.
12. *Weidlich W., Haah G.* Interregional Migration. Wien, 1988.
13. *Терехина АЮ.* Анализ данных методами многомерного шкалирования. М.: Наука, 1986.