



Проблема распознавания событий в поле центральных сил

Трунев А.П. – к. ф.-м. н., Ph.D.

Обсуждается проблема распознавания событий в поле центральных сил.

Ключевые слова: АСТРОНОМИЯ, АСТРОСОЦИОТИПОЛОГИЯ, ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ, СЕМАНТИЧЕСКИЕ ИНФОРМАЦИОННЫЕ МОДЕЛИ, СОЛНЕЧНАЯ СИСТЕМА, ГРАВИТАЦИОННОЕ ПОЛЕ.

Введение

В работах /1-2/ была сформулирована теорема астросоциотипологии, которая устанавливает зависимость функции распределения случайных событий, происходящих на земле от кинематических и динамических параметров нашей планеты при ее движении вокруг Солнца. Одним из следствий этой теоремы является то, что любые непрерывные во времени распределения событий при их представлении в зависимости от координат небесных тел содержат когерентные колебания, обусловленные движением небесных тел /3/. В работе /4/ высказана гипотеза о том, что когерентные колебания могут быть использованы для распознавания событий в астросоциотипологии. В настоящей работе обсуждается проблема распознавания событий в поле центральных сил.

Теорема астросоциотипологии

При выводе этой теоремы предполагается /1-3/, что на планете происходит ряд однородных событий, число которых в единицу времени описывается функцией $W(t)$, нормированной на единицу за один период обращения планеты вокруг центрального светила, т.е.

The events recognition problem in the field of central forces

Alexander Trunev , Ph.D.

The events recognition problem in the field of central forces is discussed in the paper.

Key words: ASTRONOMY, ASTROSOCIOTYPOLOGY, COMPUTATIONAL EXPERIMENT, SEMANTIC INFORMATION MODELS, SOLAR SYSTEM, GRAVITATION FIELD.

$$\frac{1}{T} \int_0^T W(t) dt = 1 \quad (1)$$

Период обращения связан с угловой скоростью движения по орбите интегральным соотношением

$$\frac{1}{2\pi} \int_0^T \omega(t) dt = 1$$

Для замкнутых траекторий эта теорема является следствием дифференциального уравнения, связывающего плотность функции распределения вдоль радиальной и угловой координаты в полярной системе координат

$$W(t) dt = W(t) \frac{dt}{dr} dr = W(t) \frac{dr}{d\vartheta} \frac{d\vartheta}{\dot{r}} = \frac{W(t)}{\omega(t)} d\vartheta \quad (2)$$

Отметим связь кинематических параметров:

$$\dot{r} = \omega \frac{dr}{d\vartheta} \quad (3)$$

Как известно, в поле центральных сил тело совершает финитное движение по эллипсу, уравнение которого в полярной системе координат можно представить в виде:

$$\frac{1}{r} = a + b \cos \vartheta$$

$$a = \frac{1}{2} \left(\frac{1}{r_{\min}} + \frac{1}{r_{\max}} \right), \quad b = \frac{1}{2} \left(\frac{1}{r_{\min}} - \frac{1}{r_{\max}} \right) \quad (4)$$

Здесь r_{\min}, r_{\max} - минимальное и максимальное удаление планеты от центра масс системы.

В поле центральных сил выполняется закон сохранения момента импульса в форме

$$\omega r^2 = l = \text{const}$$

Отсюда и из уравнений (3-4) вытекают простые соотношения

$$\begin{aligned}
 \dot{r} &= lb \sin \mathcal{G} \\
 \sin \mathcal{G} &= \pm \sqrt{1 - \left(\frac{1}{br} - \frac{a}{b} \right)^2} \\
 \omega(\mathcal{G}) &= l(a + b \cos \mathcal{G})^2
 \end{aligned} \tag{5}$$

В случае равновероятных событий положим в уравнениях (2) $W(t)=1$, и, используя (5), находим плотности распределения событий вдоль угловой и радиальной координаты

$$\begin{aligned}
 w(\mathcal{G}) &= c_1 / \omega(\mathcal{G}) = c_1 / l(a + b \cos \mathcal{G})^2 \\
 w(r) &= c_2 / |\dot{r}| = c_2 / lb \sqrt{1 - \left(\frac{1}{br} - \frac{a}{b} \right)^2}
 \end{aligned} \tag{6}$$

Здесь c_1, c_2 – постоянные множители, которые можно найти из условия нормирования. Отметим, что полученные плотности (6) зависят от полярного угла (или долготы Солнца) и радиальной координаты, хотя исходное распределение не зависит от времени. Плотность функции распределения в зависимости от расстояния имеет особенности в точках остановки, где радиальная скорость обращается в нуль.

В дискретном случае, рассмотренном в /3-9/ и других работах по астросоциотипологии, вместо уравнения (2) используются нормированные частоты и их стандартные отклонения:

$$\begin{aligned}
 \bar{w}_i(\mathcal{G}_j) &= \frac{N_{ij} / \sum_j N_{ij}}{\sum_i N_{ij} / \sum_{i,j} N_{ij}}, \quad N_{ij} = W_i(\mathcal{G}_j) / \omega(\mathcal{G}_j) \\
 \bar{w}_i(r_j) &= \frac{N_{ij} / \sum_j N_{ij}}{\sum_i N_{ij} / \sum_{i,j} N_{ij}}, \quad N_{ij} = \frac{W_i(r_j)}{|\dot{r}(r_j)|} \\
 \delta w_j &= \sqrt{\frac{1}{n} \sum_i \left(\bar{w}_{ij} - \frac{1}{n} \sum_i \bar{w}_{ij} \right)^2}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m
 \end{aligned} \tag{7}$$

Теорема астросоциотипологии непосредственно следует из определений (7), которые не содержат никаких внешних параметров, кроме угловой и радиальной скорости.

Отметим, что при наблюдении с земли за небесным телом, орбита которого близка к окружности, радиальная скорость в системе Земля – небесное тело связана с разностью долгот соотношением /2/:

$$\dot{r} = r_e (\omega_e - \omega_h) \sin(\mathcal{G} - \mathcal{G}_e) \quad (8)$$

Здесь $r_e, \omega_e, \mathcal{G}_e$ – радиус, угловая скорость Земли, и долгота Солнца соответственно, ω_h, \mathcal{G} – угловая скорость вращения небесного тела по орбите вокруг Солнца и его долгота соответственно.

Предположим, что известна функция распределения событий вдоль угловой координаты, $w = w(\mathcal{G})$, которую нормируем на единицу

$$\frac{1}{2\pi} \int_0^{2\pi} w(\mathcal{G}) d\mathcal{G} = 1 \quad (9)$$

Функция плотности распределения вдоль радиальной координаты может быть получена из дифференциального соотношения:

$$w(\mathcal{G}) d\mathcal{G} = w(\mathcal{G}(r)) \frac{d\mathcal{G}}{dr} dr = \tilde{w}(r) dr \quad (10)$$

Отметим, что функция плотности распределения отличается от функции с тильдой в правой части (10) на постоянный множитель, который определяется из условия нормировки. Вместо радиальной переменной удобно использовать нормированную переменную

$$x = \frac{r_{\max} - r}{r_{\max} - r_{\min}}$$

Тогда уравнение (10) приобретает вид

$$w(\mathcal{G}) d\mathcal{G} = w(\mathcal{G}(x)) \frac{d\mathcal{G}}{dx} dx = \tilde{w}(x) dx \quad (11)$$

Отсюда следует, что отношение плотностей вдоль угловой и нормированной координаты (с учетом нормировки и положительной определенности плотности) в случае Солнца равно

$$w(\mathcal{G}) / w(x) = \frac{\pi}{2} \sin \mathcal{G}, \quad 0 \leq \mathcal{G} \leq \pi \quad (12)$$

В случае произвольного небесного тела, используя уравнение (8), находим:

$$w(\mathcal{G})/w(x) \approx C|\sin(\mathcal{G} - \mathcal{G}_e)| \quad (13)$$

Здесь C – постоянный множитель, который определяется из условия нормировки.

Отметим, что уравнение (12) выполняется с точностью до величины эксцентриситета земной орбиты (или орбиты небесного тела). С той же точностью можно связать между собой функцию плотности распределения событий по времени $W(t)$ и функцию $w = w(\mathcal{G})$. Действительно, используя уравнение (2), находим

$$W(t)dt = \frac{W(t)}{\omega(t)} d\mathcal{G} \approx \frac{T}{2\pi} W(t) d\mathcal{G}$$

Таким образом, если использовать вместо истинной угловой скорости ее среднее значение, определяемое по периоду обращения планеты, тогда плотность распределения событий по углу связана с плотностью распределения событий по времени простым соотношением:

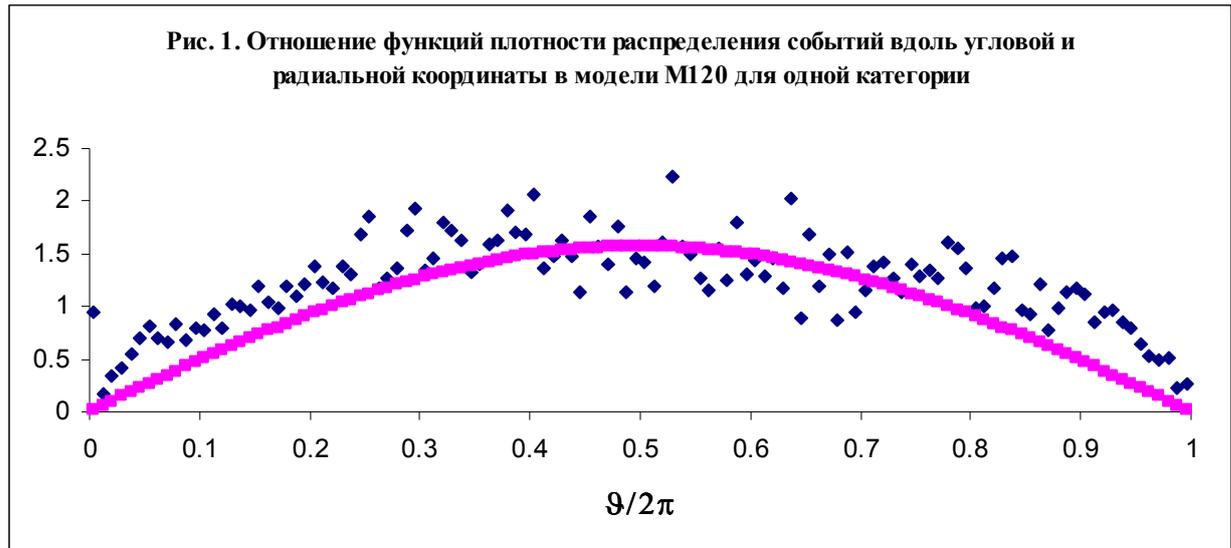
$$w(\mathcal{G}) \approx \frac{T}{2\pi} W\left(\frac{T\mathcal{G}}{2\pi}\right) \quad (14)$$

Далее заметим, что если функция $W(t)$ является регулярной, то и функция $w = w(\mathcal{G})$ является регулярной в силу (14). Но тогда функция $w = w(x)$ может иметь особенности в точках, где $\sin \mathcal{G} = 0$, в силу уравнения (13) или в точках, где $\sin(\mathcal{G} - \mathcal{G}_e) = 0$, в силу уравнения (13). Очевидно, что эти свойства относятся к непрерывным распределениям событий, тогда как в случае дискретных событий особенности заменяются конечными величинами – пиками событий /1-2/.

Можно показать, что в случае дискретных распределений уравнение (12) имеет дискретный аналог

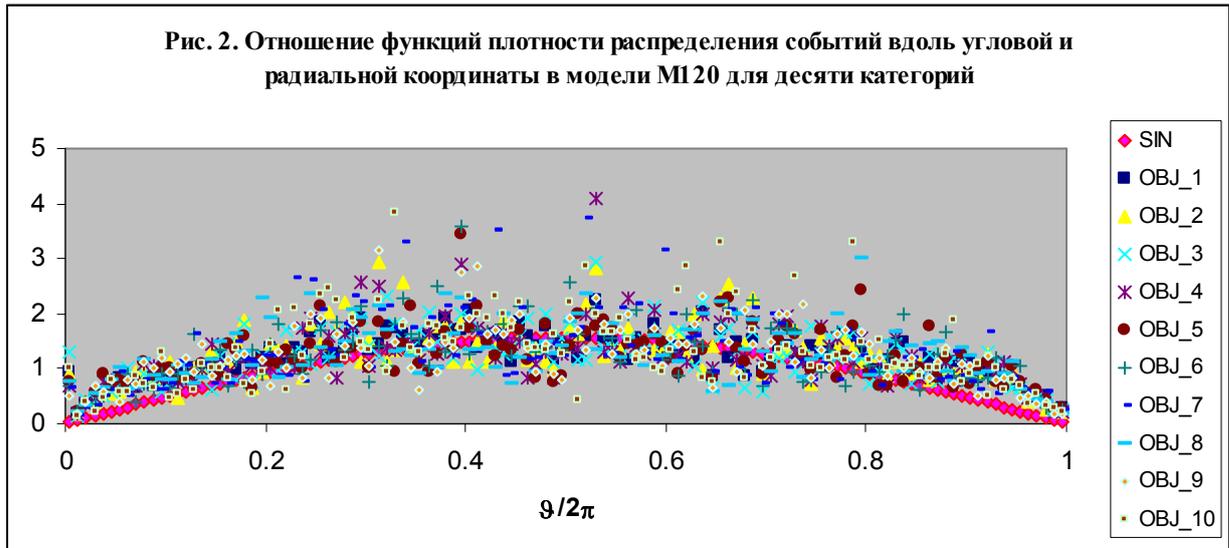
$$N_{ij}(\mathcal{G}_j)/N_{ij}(x_j) \cong \frac{\pi}{2} \sin \mathcal{G}_j, \quad 0 \leq \mathcal{G}_j \leq \pi \quad (15)$$

Уравнение (15) выполняется тем точнее, чем больше общее число случаев и число ячеек модели, а также число случаев, приходящихся на одну ячейку.

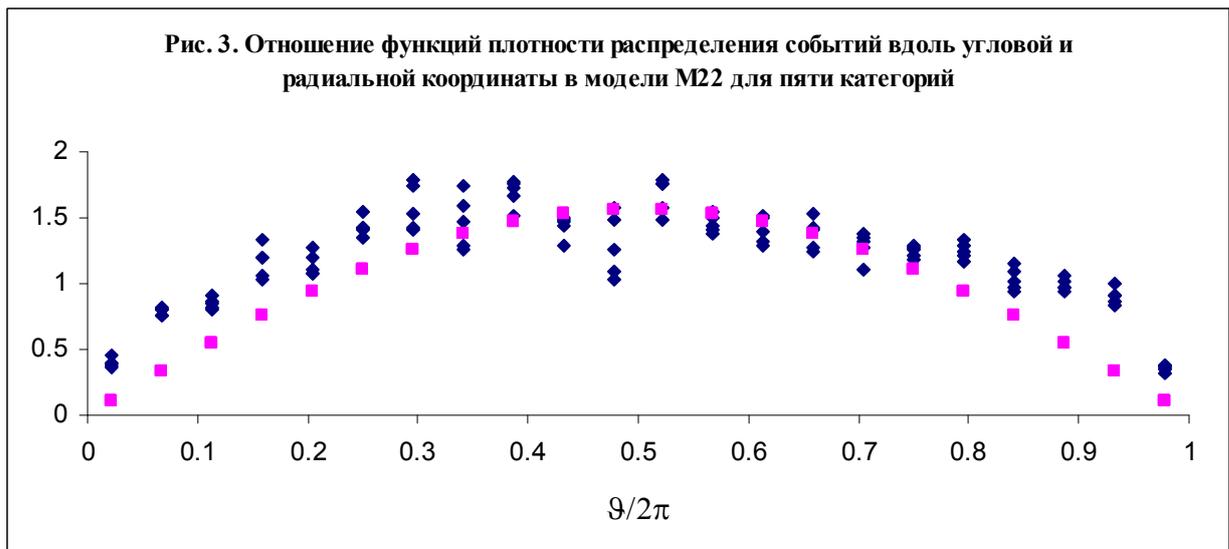


На рис. 1 представлены данные отношения распределений $N_{ij}(\vartheta_j)/N_{ij}(x_j)$ реализации выбора 20007 респондентов из 37 категорий для одной категории с общим числом случаев 13640 в зависимости от нормированного угла $\vartheta_j/2\pi$ в модели M120. Числа $N_{ij}(\vartheta_j), N_{ij}(x_j)$ определялись на основе комплекса программ «Эйдос-астра» /10-11/ путем суммирования записей банка данных, относящихся к данной категории в данной ячейке. Сплошная линия на рис. 1 представляет правую часть уравнения (15). Наблюдается линейная корреляция данных с теоретической кривой с коэффициентом $R^2 = 0.6443$.

Менее представительные категории имеют большее рассеяние данных относительно теоретической кривой. Поэтому по мере роста числа категорий рассеяние данных возрастает. Тем не менее, данные всех 37 исследованных категорий ложатся вокруг теоретической кривой. На рис. 2 представлены данные отношения распределений $N_{ij}(\vartheta_j)/N_{ij}(x_j)$ реализации выбора 20007 респондентов из 37 категорий для десяти наиболее представительных категорий из таблицы 1 работы /9/ с общим числом случаев 44176 в зависимости от нормированного угла $\vartheta_j/2\pi$, в модели M120. Сравнивая данные на рис. 1 и 2 можно сделать вывод, что рассеяние данных относительно теоретической кривой (15) заметно растет с ростом числа категорий. Это связано с тем, что при уменьшении общего числа случаев приходящихся на категорию, уменьшается и число случаев, приходящихся на одну ячейку модели.



Уменьшая число ячеек модели, можно понизить рассеяние данных относительно теоретической кривой за счет увеличения числа случаев, приходящихся на одну ячейку. На рис. 3 представлены данные отношения распределений $N_{ij}(\vartheta_j) / N_{ij}(x_j)$ реализации выбора 20007 респондентов из 37 категорий для пяти наиболее представительных категорий с общим числом случаев 31176 в зависимости от нормированного угла $\vartheta_j / 2\pi$ в модели M22. Для совокупности этих данных наблюдается линейная корреляция с теоретической кривой (представлена квадратными символами на рис. 3) с коэффициентом $R^2 = 0.7928$.



Таким образом, данные, представленные на рис. 1-3, свидетельствуют, что дискретный аналог уравнения (12) в форме (15) действительно существует. Докажем это утверждение, используя инте-

гральную форму выражения числа событий в данной ячейке и теорему о средней точке, имеем:

$$N_{ij}(\vartheta_j) = N_i \int_{\vartheta_j}^{\vartheta_j + \Delta\vartheta} w(\vartheta) d\vartheta = N_i w(\tilde{\vartheta}_j) \Delta\vartheta, \quad \vartheta_j < \tilde{\vartheta}_j < \vartheta_j + \Delta\vartheta$$

$$N_{ij}(x_j) = N_i \int_{x_j}^{x_j + \Delta x} w(x) dx = N_i w(\tilde{x}_j) \Delta x, \quad x_j < \tilde{x}_j < x_j + \Delta x$$

Здесь N_i – общее число случаев данного типа. С учетом полученных выражений составим отношение

$$\frac{N_{ij}(\vartheta_j)}{N_{ij}(x_j)} = \frac{w(\tilde{\vartheta}_j) \Delta\vartheta}{w(\tilde{x}_j) \Delta x}$$

Далее заметим, что в дискретном случае при выборе равномерной сетки по угловой и радиальной координате, как в работах /1-7/, отношение $\Delta\vartheta / \Delta x = const$. Без ограничения общности можно выбрать эту константу так, чтобы получить выражение (15). Следовательно, уравнение (12) действительно выполняется в дискретном случае, причем невязка, необходимая для согласования правой и левой части уравнения (15), определяется выбором средней точки в соответствующих интервалах, т.е.

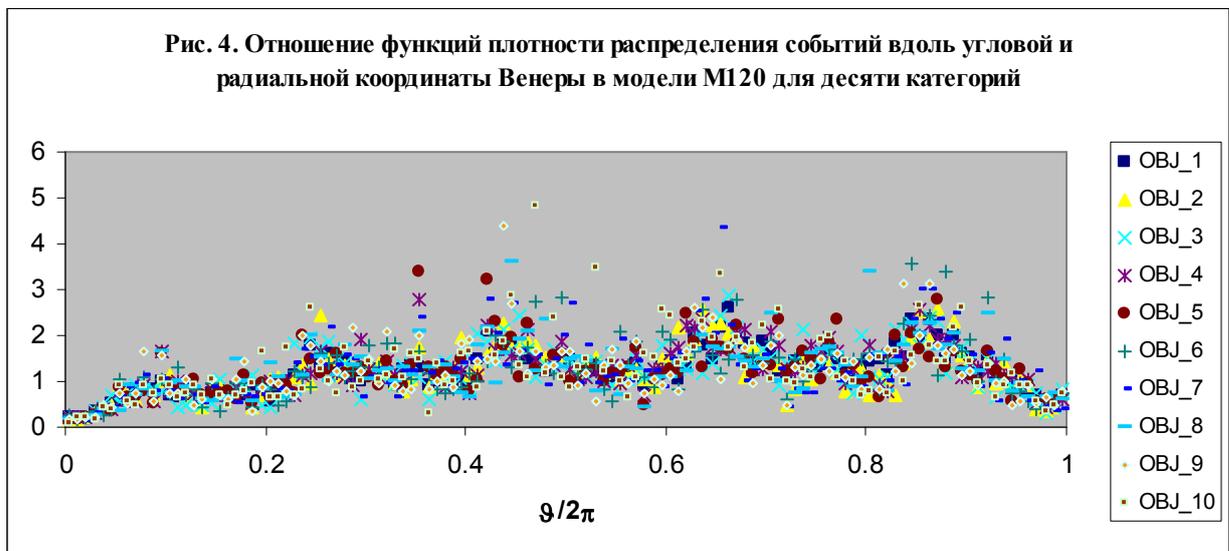
$$\frac{N_{ij}(\vartheta_j)}{N_{ij}(x_j)} = \frac{\pi}{2} \sin \vartheta_j + \frac{w(\tilde{\vartheta}_j)}{w(\tilde{x}_j)} - \frac{w(\vartheta_j)}{w(x_j)} \quad (16)$$

Для дифференцируемой функции плотности распределения легко показать, что невязка в правой части уравнения (16) стремится к нулю при условии, что $\Delta\vartheta, \Delta x \rightarrow 0$. Однако, для дискретных распределений это выполняется лишь в том случае, если число случаев, приходящихся на одну ячейку, достаточно велико, что заведомо не выполняется в задачах с конечным числом событий. Для таких задач можно оптимизировать невязку, путем перебора числа ячеек сетки. В частности, для данных, приведенных на рис. 1-3, оптимальное число ячеек сетки $M=22$.

Полученные выше результаты касаются свойств функции плотности распределения событий по угловой и радиальной координате

Солнца относительно нашей планеты. К сожалению, для других небесных тел солнечной системы не удастся получить простой дискретный аналог уравнения (13), поскольку в правую часть этого уравнения входит разность углов, а не один угол, как в уравнении (12). Кроме того, даже если исходное распределение событий $W(t)$ является случайным, распределения событий вдоль долготы и расстояния до небесных тел содержат когерентную составляющую, обусловленную периодическим движением нашей планеты.

На рис. 4 представлены данные отношения распределений по долготы и расстоянию до Венеры - $N_{ij}(\vartheta_j) / N_{ij}(x_j)$, реализации выбора 20007 респондентов из 37 категорий для десяти наиболее представительных категорий из таблицы 1 работы /9/ с общим числом случаев 44176 в зависимости от нормированного угла $x = \vartheta_j / 2\pi$ в модели M120. В распределении данных отчетливо наблюдаются когерентные колебания, обусловленные периодическим движением Земли и Венеры.

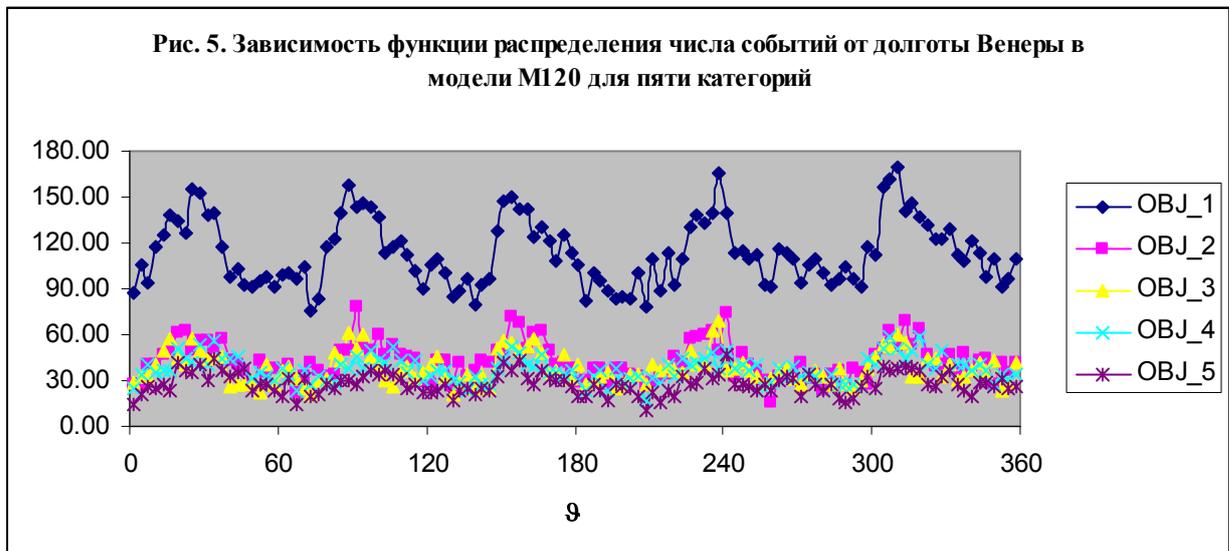


На рис. 5 представлены данные абсолютных распределений $N_{ij}(\vartheta_j)$ реализации выбора 20007 респондентов из 37 категорий для пяти наиболее представительных категорий с общим числом случаев 31176 в зависимости от долготы Венеры в модели M120. Из этих данных следует, что распределение событий по долготе Венеры содержит колебания, обусловленные изменением угловой скорости в системе Венера-Земля. Действительно, как это следует из уравне-

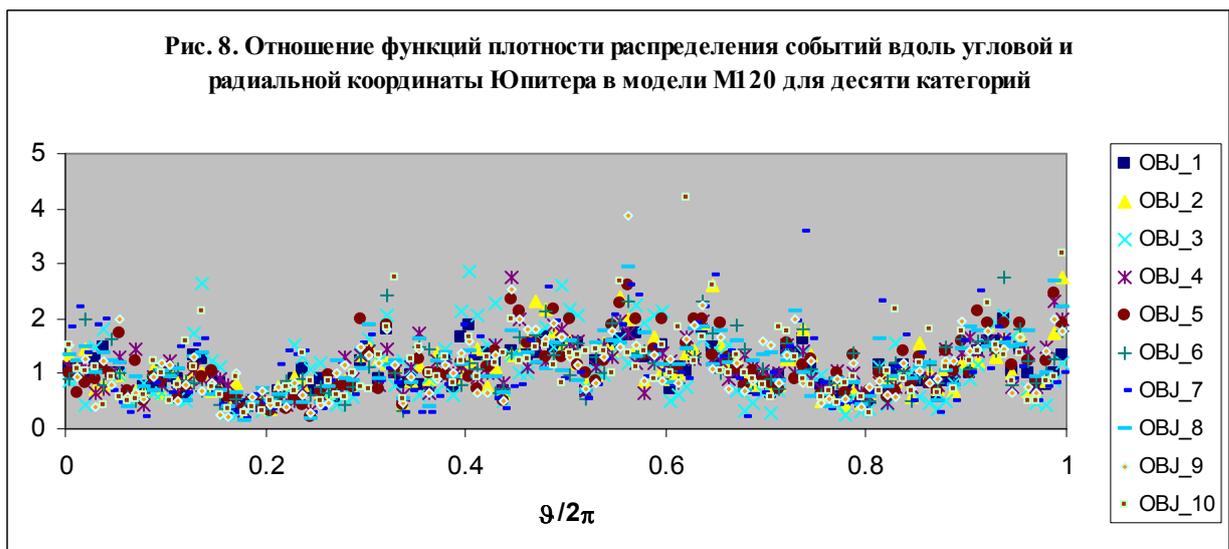
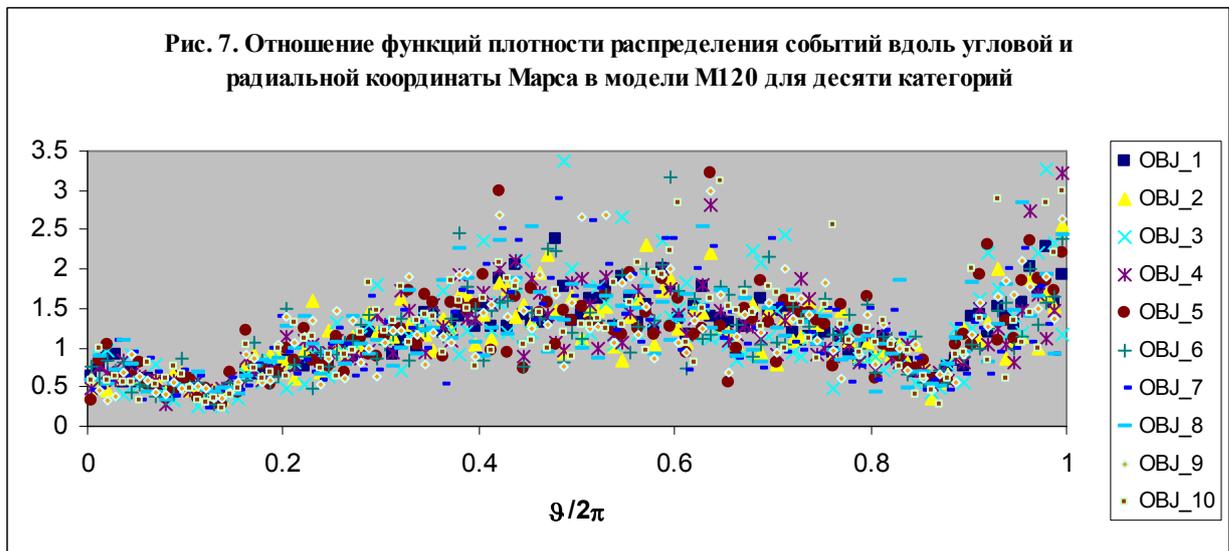
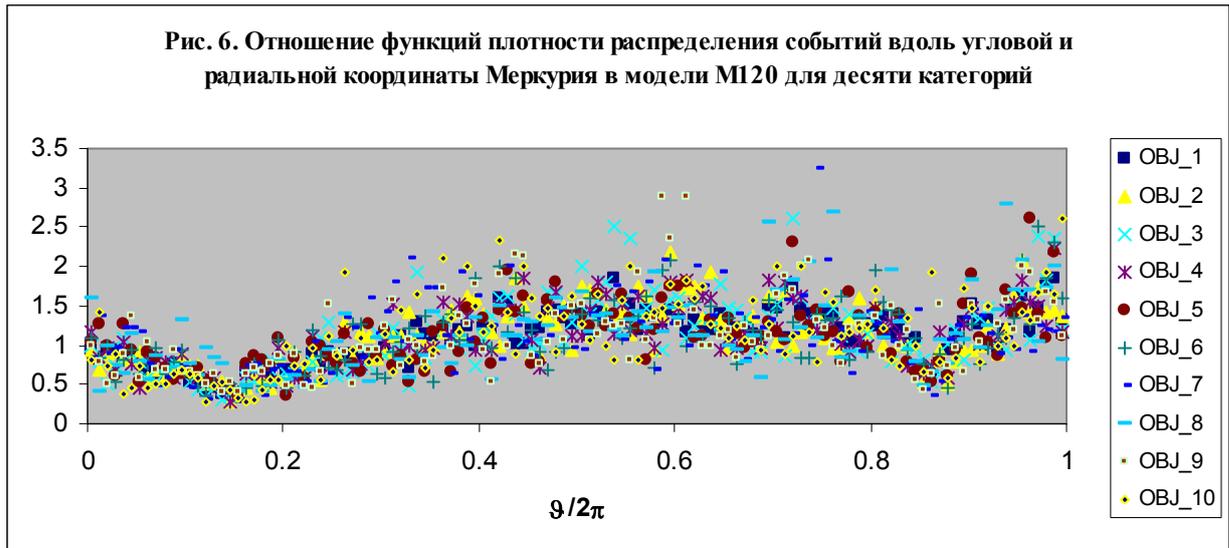
ния (2), плотность распределения событий вдоль угловой координаты изменяется обратно пропорционально угловой скорости, т.е.

$$w(\vartheta)d\vartheta = \frac{W(t)}{\omega(t)}d\vartheta \quad (17)$$

В случае Солнца это изменение невелико и дается первым уравнением (6). В случае других небесных тел относительное движение велико, что приводит к большому изменению относительной угловой скорости и к появлению колебаний в распределении числа событий, как это видно из данных на рис. 5.

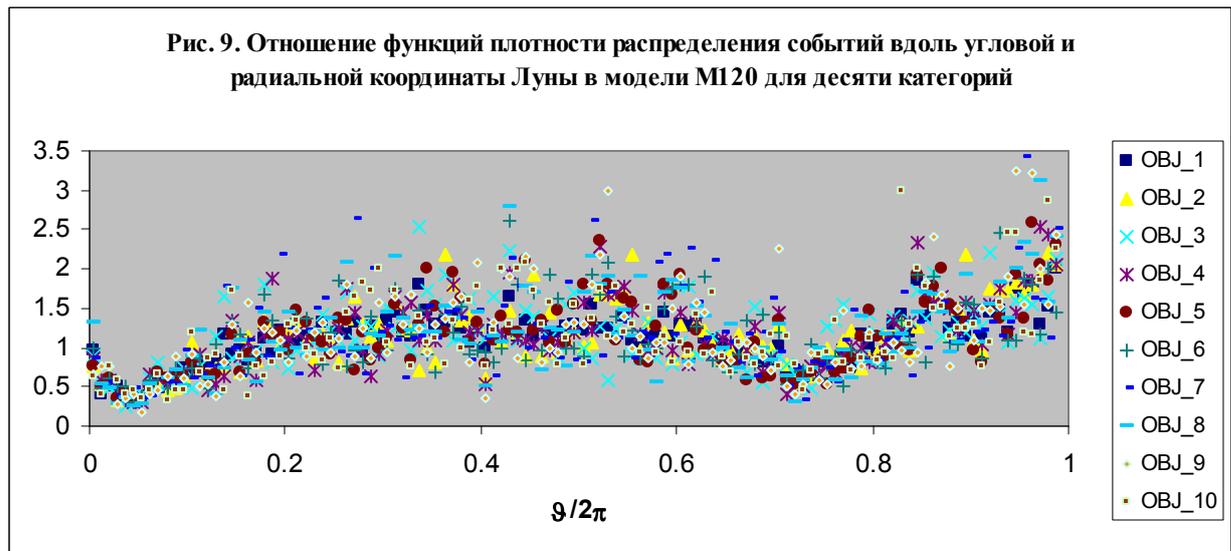


Заметим, что указанным свойством квазипериодичности, обусловленным изменением знака относительной угловой скорости, обладают и распределения относительно долготы других небесных тел. При этом, влияние каждого небесного тела на функцию распределения можно достаточно легко идентифицировать по внешнему виду кривой отношения функций плотности распределения вдоль угловой и радиальной координаты. На рис. 6-9 представлены данные отношения распределений по долготе и расстоянию - $N_{ij}(\vartheta_j)/N_{ij}(x_j)$, реализации выбора 20007 респондентов из 37 категорий для десяти наиболее представительных категорий из таблицы 1 работы /9/ с общим числом случаев 44176 в зависимости от нормированного угла $x_j = \vartheta_j / 2\pi$ в модели M120 для Меркурия, Марса Юпитера и Луны соответственно. Можно видеть, что данные на рис. 6-9 группируются сходным образом, но иначе, чем данные для Солнца (рис. 2) и Венеры (рис.4).



Данные для Юпитера на рис. 8 содержат 12 пиков, обусловленных взаимным когерентным движением Земли и Юпитера на протяжении

нии 12-летнего цикла. Данные для Марса и Меркурия при внешнем сходстве рис. 6 и 7 имеют разное рассеяние относительно среднего значения. Данные для Луны группируются аналогично данным для Меркурия и Марса, но со своими специфическими особенностями. Отметим, что все данные на рис. 2, 4, 6-9 изменяются пропорционально радиальной скорости в системе Земля – соответствующее небесное тело.



Таким образом, в силу уравнения (2), данные распределений относительно координат любого небесного тела – расстояния и долготы, содержат когерентные составляющие, обусловленные периодическим движением небесного тела относительно Земли. Зададимся вопросом, можно ли использовать когерентные возмущения функции плотности распределения данных, обусловленные движением небесных тел для распознавания самих данных по астрономическим параметрам, заданным в момент события $W(t)$? Ответ является положительным в задачах астросоциотипологии, связанных с распознаванием социальных категорий, причем можно утверждать, что чем больше использовано разнородных входных астрономических параметров, тем выше вероятность распознавания. Например, сочетание параметров долготы и расстояния должно приводить к повышению вероятности распознавания, что было обнаружено в работе /6/. С другой стороны, при замене независимых от долготы небесных тел углов домов на зависимые от долготы параметры склонения, вероятность распознавания должна снижаться, что было обнаружено в работе /14/.

Задача о распознавании социальных категорий в поле центральных сил

Сформулируем задачу распознавания социальных категорий, рассматривая их как совокупность событий выбора самих категорий множеством респондентов. Тогда момент рождения респондентов можно отнести к началу события. Действительно, выбор социальной категории, например, профессии, не является одномоментным событием. Обычно ему предшествуют периоды воспитания и образования, практической подготовки. Однако для упрощения задачи удобно отнести это событие ко времени рождения респондентов, поскольку в этой задаче нет никакого другого выделенного момента времени, кроме рождения индивида.

Итак, имеется множество респондентов A , которому ставится в соответствие множество категорий C_i . Каждый респондент характеризуется моментом времени и местом рождения (географические координаты места рождения). По данным рождения можно построить матрицу, содержащую координаты небесных тел, например углы долготы и расстояния, а по биографическим данным респондентов можно определить частотные распределения N_i – число респондентов, имеющих отношение к данной категории C_i . В частности, для банка данных /12/ соответствующие распределения приведены в работах /5-9/ и других. Отметим, что один респондент может иметь отношение к нескольким категориям.

Определим число случаев реализации данной категории, которое приходится на заданный интервал изменения астрономических параметров, имеем в дискретном случае:

$$\begin{aligned}
 N_{ij}(\mathcal{G}_j, k) &= N_i w(\tilde{\mathcal{G}}_j, k) \Delta \mathcal{G}, & \mathcal{G}_j < \tilde{\mathcal{G}}_j < \mathcal{G}_j + \Delta \mathcal{G} \\
 N_{ij}(x_j, k) &= N_i w(\tilde{x}_j, k) \Delta x, & x_j < \tilde{x}_j < x_j + \Delta x \\
 1 \leq i \leq n, & 1 \leq j \leq m, & k = 1, \dots, k_0
 \end{aligned} \tag{18}$$

Здесь k_0 – число небесных тел, используемых в задаче.

В системе искусственного интеллекта /10/, используемой для решения задач астросоциотипологии, осуществляется фильтрация сигнала путем определения матрицы информативности в виде /13/

$$\begin{aligned}
I_{ijk} &= \log_2 \frac{N_{ij} / \sum_i N_{ij}}{\sum_i N_{ij} / \sum_{i,j} N_{ij}}, \quad N_{ij}(x_{jk}) \neq 0 \\
I_{ijk} &= 0, \quad N_{ij}(x_{jk}) = 0, \\
\delta I_{jk} &= \sqrt{\frac{1}{n} \sum_i \left(I_{ijk} - \frac{1}{n} \sum_i I_{ijk} \right)^2} \\
1 \leq i \leq n, \quad 1 \leq j \leq 2m, \quad 1 \leq k \leq k_0
\end{aligned} \tag{19}$$

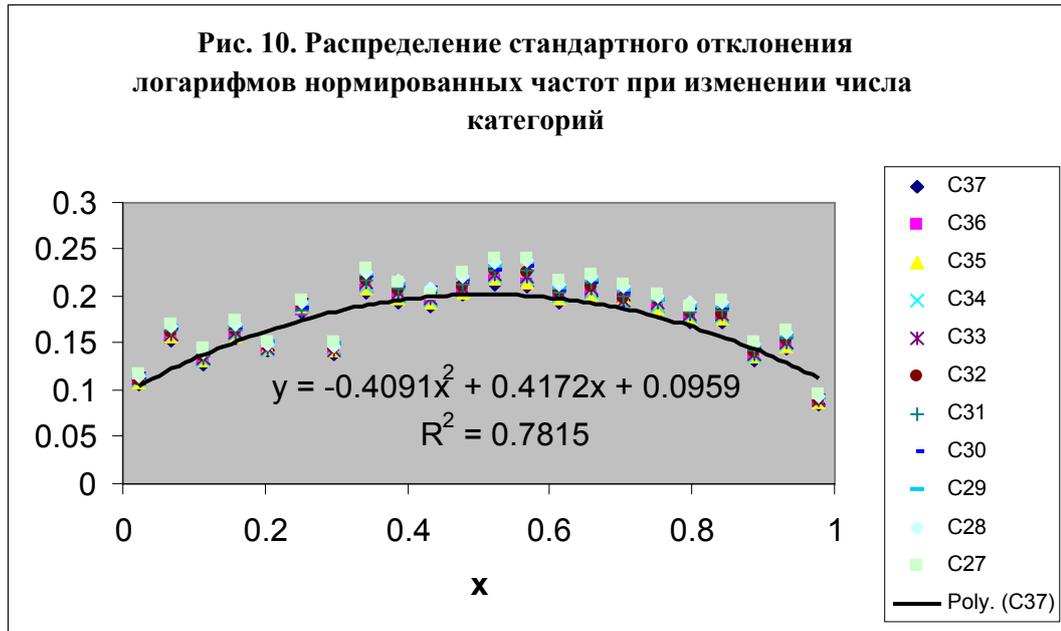
Здесь нормированная переменная определена следующим образом:

$$x_{jk} = \begin{cases} \mathcal{G}_j(k) / 2\pi, & 1 \leq j \leq m \\ \frac{r_{\max}(k) - r(k)}{r_{\max}(k) - r_{\min}(k)}, & m + 1 \leq j \leq 2m \end{cases}$$

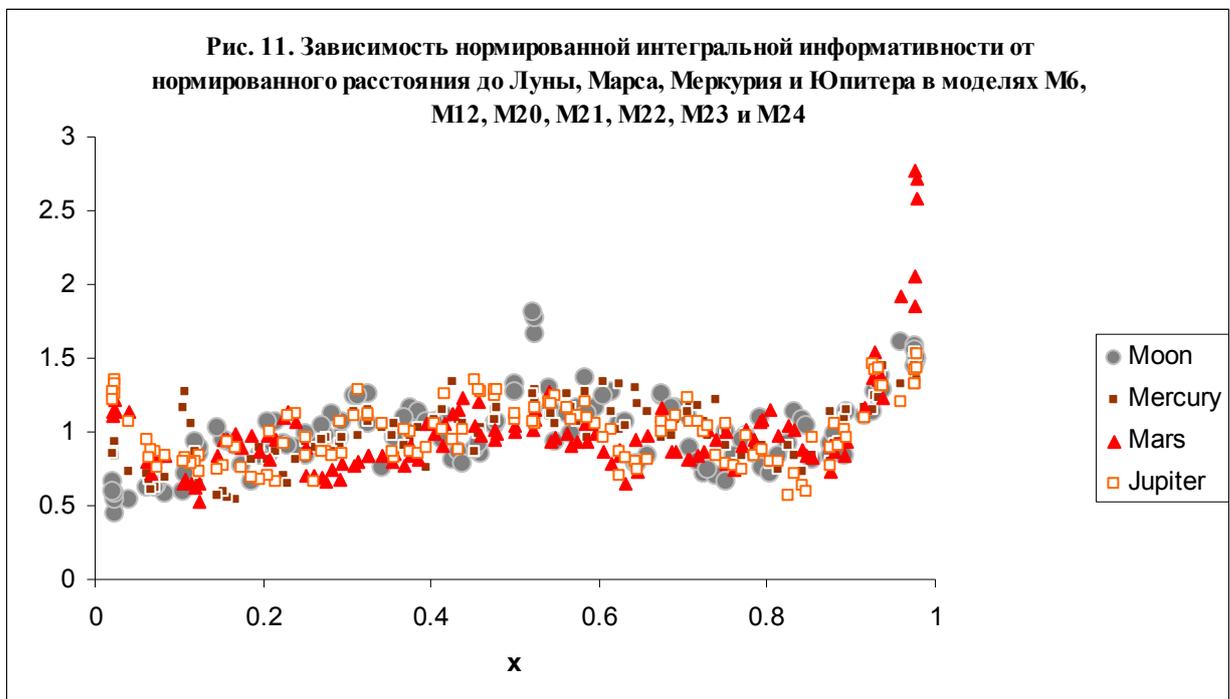
Первая величина (19) называется информативность признака, а вторая величина является стандартным отклонением информативности или интегральная информативность (ИИ).

В случае непрерывных распределений событий определенная согласно (19) информативность явно не зависит от кинематических параметров небесных тел. Однако в дискретных задачах было обнаружено, что ИИ зависит от расстояния до небесных тел /5-9/, что обусловлено невязкой, возникающей в формулах типа (16).

На рис. 10 представлены результаты расчетов интегральной информативности в зависимости от расстояния до Солнца в модели M22 выбора из 27-37 категорий (маркеры C27-C37 соответственно) для 20007 респондентов из банка данных /12/. Отметим, что зависимость типа представленной на рис. 10 наблюдается в широком диапазоне изменения числа интервалов разбиения от 6 до 172, числа категорий от 11 до 100 и общего числа случаев от 13479 до 123671 /8-9/. Таким образом, эту закономерность можно признать фундаментальным фактом. По характеру распределения данных на рис. 6 можно сделать вывод, что ИИ изменяется пропорционально радиальной скорости, что непосредственно следует из основной теоремы астросоциотипологии /1-3/.



На рис. 11 представлены обобщенные данные нормированного параметра ИИ в зависимости от нормированного расстояния до Луны, Марса, Меркурия и Юпитера, полученные на сетках семи масштабов. Сравнивая данные, представленные на рис. 11 и на рис. 6-9, можно сделать вывод, что для этих небесных тел ИИ изменяется пропорционально радиальной скорости, как и в случае Солнца.



Следовательно, зависимость статистических параметров от кинематических параметров небесных тел сохраняется в матрице информативности, определенной согласно первому уравнению (19). Определим размерность этой матрицы для примеров, приведенных на

рис. 1-9. В модели M120 матрица информативности содержит $37 \times 120 \times 20 = 88800$ элементов для 37 категорий и 20 параметров десяти небесных тел, разбитых на 120 ячеек. Соответственно в модели M22 матрица информативности содержит $37 \times 22 \times 20 = 16280$ элементов.

Каждой категории можно сопоставить вектор информативности астрономических параметров размерности $2mk_0$, составленный из элементов матрицы информативности, путем последовательной записи столбцов, соответствующих нормированной координате, в один столбец, т.е.

$$c_{is} = I_{ijk} \Big|_{jk=s}, \quad 1 \leq s \leq 2mk_0 \quad (20)$$

С другой стороны, процесс идентификации и распознавания может рассматриваться как разложение вектора распознаваемого объекта в ряд по векторам категорий (классов распознавания) /13/. Этот вектор, состоящий из единиц и нулей, можно определить по координатам небесных тел, соответствующих дате и месту рождения респондента l в виде

$$a_{ls} = \begin{cases} 1, & (j-1)\Delta x \leq x_{jk}(l) \leq j\Delta x, \quad jk = s \\ 0, & 1 \leq s \leq 2mk_0 \end{cases} \quad (21)$$

Таким образом, если нормированная координата небесного тела из данных респондента попадает в заданный интервал, элементу вектора придается значение 1, а во всех остальных случаях – значение 0. Перечисление координат осуществляется последовательно, для каждого небесного тела.

В случае, когда система векторов (20) является полной, можно любой вектор (21) представить в виде линейной комбинации векторов системы (20). Коэффициенты этого разложения будут соответствовать уровню сходства данного респондента с данной категорией. В случае неполной системы векторов (20) точная процедура заменяется распознаванием. При этом уровень сходства данных респондента с той или иной категорией можно определить по величине скалярного произведения вектора (21) на вектор (20), т.е.

$$K_{il} = \frac{1}{|a_l||c_i|} \sum_{s=1}^{2mk_0} a_{ls}(A)c_{is} \quad (22)$$

Отметим, что возможны четыре исхода, при которых можно истинно или ложно отнести или не отнести данного респондента к данной категории. Для учета этих исходов распознавание социальных категорий в системе «Эйдос-астра» /10/ осуществляется по параметру сходства, который определяется следующим образом /15/:

$$S_i = \frac{1}{N} \sum_{l=1}^N (BT_{il} + T_{il} - BF_{il} - F_{il}) \cdot 100 \% \quad (23)$$

S_i – достоверность идентификации « i -й» категории;

N – количество респондентов в распознаваемой выборке;

BT_{il} – уровень сходства « l -го» респондента с « i -й» категорией, к которой он был правильно отнесен системой;

T_{il} – уровень сходства « l -го» респондента с « i -й» категорией, к которой он был правильно не отнесен системой;

BF_{il} – уровень сходства « l -го» респондента с « i -й» категорией, к которой он был ошибочно отнесен системой;

F_{il} – уровень сходства « l -го» респондента с « i -й» категорией, к которой он был ошибочно не отнесен системой.

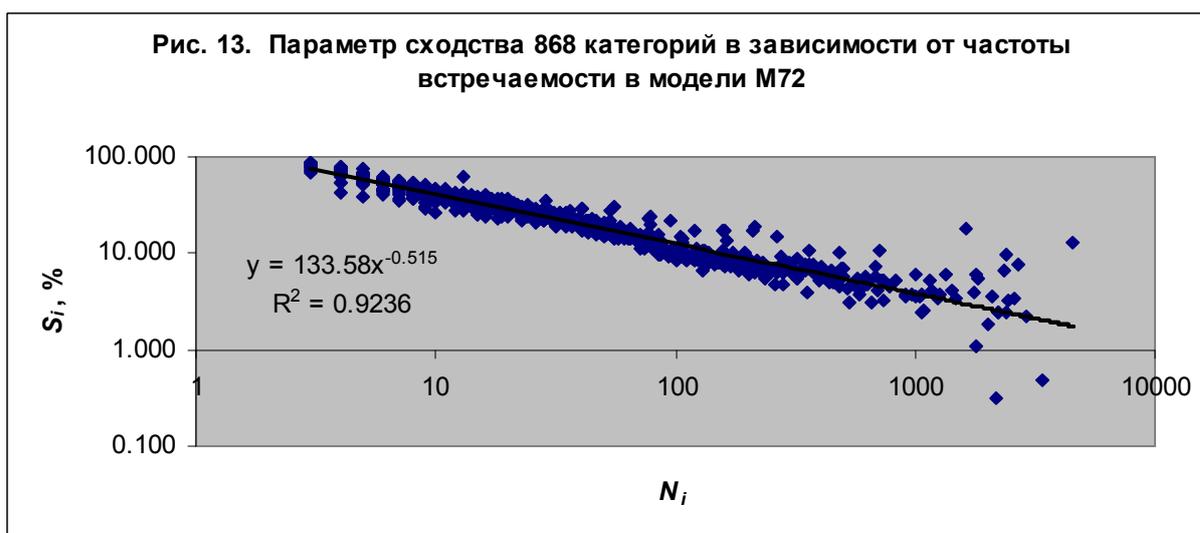
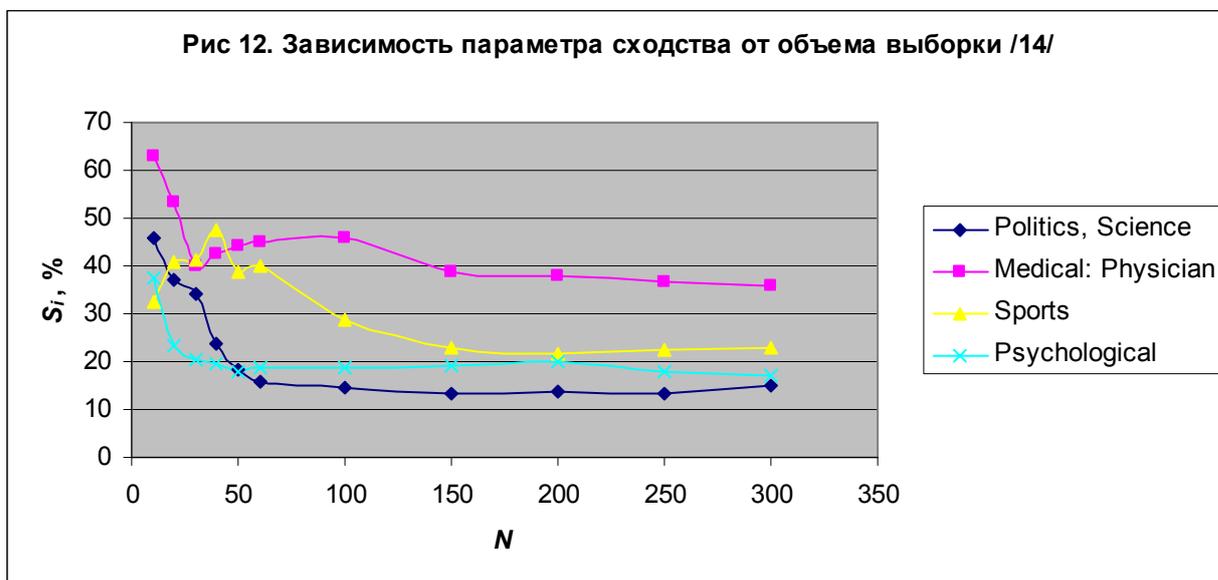
При таком определении параметр сходства изменяется в пределах от -100% до 100%, как обычный коэффициент корреляции в статистике. На рис. 12 представлен параметр сходства четырех категорий при изменении объема выборки, взятый из работы /14/. Можно отметить, что уже при $N > 150$ происходит стабилизация параметра сходства различных категорий. Очевидно, что параметр сходства должен удовлетворять критерию простой проверки

$$S_i(N_i = 1) = 100 \%$$

На рис 13 представлена зависимость параметра сходства 868 категорий от частоты встречаемости в модели M72. В этом примере минимальная частота равна 3, а критерий простой проверки выполняется, согласно уравнению степенной зависимости, при $N_i = 1.7545$, т.е. в интервале от 1 до 2. Следовательно, при малом объеме выборки сравнение данных в системе /10/ осуществляется простым перебором, а при большом объеме выборки используются статистические методы.

При заданном объеме выборки и при фиксированном числе небесных тел, параметр сходства зависит от числа ячеек модели. Как было установлено /5, 7, 15/ средний параметр сходства 37 кате-

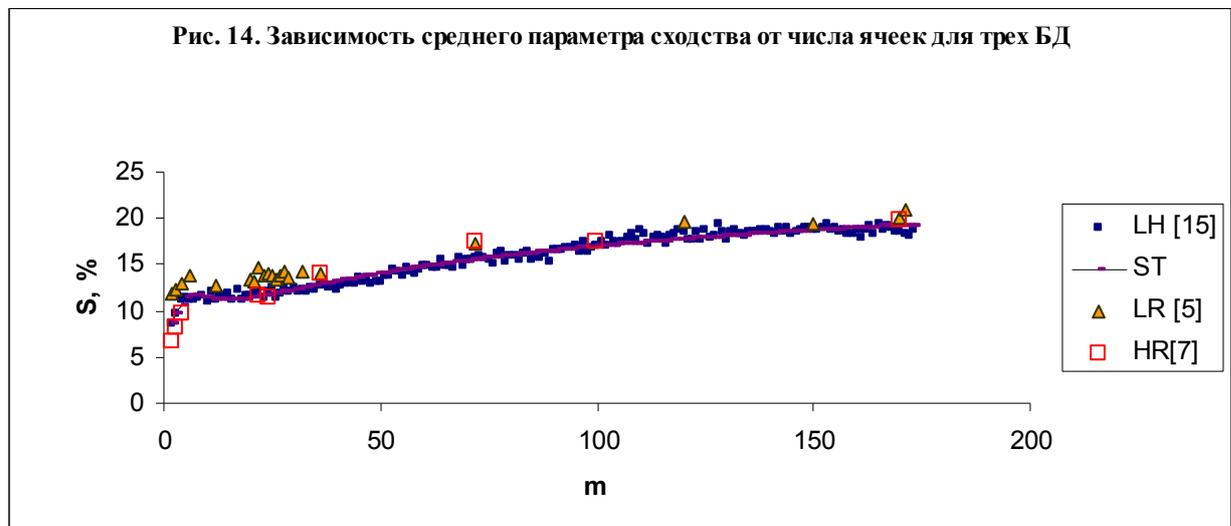
горий, данные для которых приведены на рис. 1-11, возрастает с числом ячеек модели по логарифмическому закону - см. рис. 14.



Эти результаты показывают, что процедура распознавания по параметру сходства (23), реализованная в системе «Эйдос-астра» /10/, является устойчивой как относительно объема выборки, так и относительно числа ячеек модели. Математическое обоснование этой процедуры дано в монографии /13/, где также указана связь использованной меры (19) со статистикой χ^2 .

Естественно возникает вопрос, можно ли распознать произвольного респондента, не входящего в исходное множество A ? Чтобы ответить на этот вопрос, разобьем исходное множество A на два подмножества – A_1 , A_2 . Сформируем матрицу информативности категорий, используя только подмножество A_1 . Выполним процедуру

распознавания респондентов из подмножества A_2 . Теоретически определить процент угадывания при таком распознавании можно только при случайном выборе. Практический же ответ на поставленный вопрос был получен в работе /15/. Так, например, из 5000 респондентов тестируемой выборки A_2 , чьи данные представлены на сайте /12/, в системе "Эйдос-астра" /10/ социальные категории были определены правильно у 68,75% респондентов. Обучение системы осуществлялось в этой задаче на базе данных 15007 респондентов.



Причина, по которой оказывается возможным идентифицировать подмножества (категории) событий различной даже случайной природы, используя астрономические параметры и матрицу информативности, достаточно очевидна. Ведь фактически идентифицируются распределения, которые образуются при модулировании исходных распределений астрономическими параметрами. В некоторых случаях этого достаточно, чтобы осуществить распознавание категорий.

Эту задачу можно сравнить с разложением солнечного света призмой. Свет представляет собой смесь случайных электромагнитных колебаний, но после прохождения призмы свет разлагается в спектр категорий цвета. При этом природа самого света не меняется.

Аналогично множество $W(t)$ может представлять собой смесь случайных событий, но в процессе обработки системой формируется матрица информативности астрономических параметров, которую

можно использовать для распознавания категорий. При этом случайная природа множества $W(t)$ не меняется.

Наконец заметим, что множество событий $W(t)$ может иметь любую природу, например, это могут быть ежедневные данные котировок валют разных стран /4, 11, 16/. При этом в качестве категорий выступают сами валюты, вернее, их котировки в отношении к американскому доллару. Предложенная в работе /16/ модель позволяет прогнозировать котировки валют на день вперед. В процессе моделирования был определен параметр отношения истинных прогнозов к ложным прогнозам для 20 стран – таблица 1.

Таблица 1. Качество прогноза котировок валют на день вперед для 20 стран /16/

Country	True/False	True%	False%
UK	1,142,857	53.00%	47.00%
TAIWAN	1,454,545	59.00%	41.00%
SWITZERLAND	1,545,455	61.00%	9.00%
SINGAPORE	2	67.00%	33.00%
AUSTRALIA	1	50.00%	50.00%
SWEDEN	1,071,429	52.00%	48.00%
NEW ZEALAND	0,578947	37.00%	63.00%
NORWAY	0,928571	48.00%	52.00%
MEXICO	0,928571	48.00%	52.00%
MALAYSIA	2,428,571	71.00%	29.00%
KOREA	1,9	66.00%	34.00%
JAPAN	0,733333	42.00%	58.00%
INDIA	1,363,636	58.00%	42.00%
HONG KONG	0,833333	45.00%	55.00%
EURO	1,333,333	57.00%	43.00%
DENMARK	1,357,143	58.00%	42.00%
CHINA	3,2	76.00%	24.00%
CANADA	1,307,692	57.00%	43.00%
BRAZIL	1,285,714	56.00%	44.00%
SOUTH_AFRICA	1,818,182	65.00%	35.00%

Из приведенных в этой таблице данных следует, что использование астрономических параметров позволяет делать 76 % правильных прогнозов в случае Китая и 66% в случае Кореи. Таким образом, предложенный метод можно использовать для распознавания событий любой природы, включая события, связанные с техногенными катастрофами /17/.

Автор выражает благодарность профессору Е.В. Луценко за полезные обсуждения.

Ссылки

1. Alexander Trunev. О распределении случайных событий в поле центральных сил/Chaos and Correlation, April 29 , 2009, http://trunev.com/Chaos/CR_4_2009.pdf
2. Трунев А.П. РАСПРЕДЕЛЕНИЕ СЛУЧАЙНЫХ СОБЫТИЙ В ПОЛЕ ЦЕНТРАЛЬНЫХ СИЛ/Научный журнал КубГАУ, №49(05), 2009. <http://ej.kubagro.ru/2009/05/pdf/01.pdf>
3. Alexander Trunev. Свойства непрерывных и дискретных распределений событий в поле центральных сил/Chaos and Correlation, May 24, 2009. http://trunev.com/Chaos/CR_5_2009.pdf
4. Трунев А.П. Теория и прикладные вопросы астросоциотипологии/ Научный журнал КубГАУ, №49(05), 2009. <http://ej.kubagro.ru/2009/05/pdf/04.pdf>
5. Луценко Е.В., Трунев А.П. Фундаментальные законы распознавания социальных категорий по астрономическим данным/ Chaos and Correlation, January 25, 2009, http://trunev.com/Chaos/CR1_1_2009.pdf
6. Луценко Е.В., Трунев А.П. Исследование зависимости интегральной информативности от расстояния до небесных тел Солнечной системы/ Chaos and Correlation, Dec. 8, 2008.
7. Eugene Lutsenko, Alexander Trunev. О фундаментальных закономерностях распознавания социальных категорий группы субъектов по астрономическим данным на момент рождения/Chaos and Correlation. International Journal, December 21st, 2008. http://trunev.com/Chaos/CR12_1.pdf
8. Луценко Е.В., Трунев А.П. Исследование вариабельности интегральной информативности моделей реагирования субъектов на положение небесных тел солнечной системы в момент рождения/ Научный журнал КубГАУ, № 45(1), 2009.
9. Луценко Е.В., Трунев А.П. Устойчивость зависимости интегральной информативности от расстояния до небесных тел Солнечной системы/ Научный журнал КубГАУ, № 46(2), 2009, <http://ej.kubagro.ru/2009/02/pdf/12.pdf>
10. Patent 2008610097, Russia, System for Typification and Identification of the Social Status of Respondents Based on the Astronomical Data at the Time of Birth - "AIDOS-ASTRO" / E.V. Lutsenko, A.P. Trunev, V.N. Shashin; Application № 2007613722, January 9, 2008.
11. Трунев А. П., Луценко Е. В. Астросоциотипология: Монография (научное издание). Краснодар: КубГАУ, 2008.
12. Astro-Databank Wiki Project http://www.astro.com/astro-databank/Main_Page
13. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем). Краснодар: КубГАУ, 2002, – 605 с.
14. Eugene Lutsenko, Alexander Trunev. Artificial intelligence system for identification of social categories of natives based on astronomical parameters. Научный журнал КубГАУ, № 03(37), 2008.
15. Луценко Е.В., Трунев А.П. Астросоциотипология и спектральный анализ личности по астросоциотипам с применением семантических информационных мультимodelей. Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008.– №1(35). – Режим доступа: <http://ej.kubagro.ru/2008/01/pdf/10.pdf>
16. Alexander Trunev, Victor Okhonin. About a correlation between currency exchange rate and astronomical parameters of the Solar system's celestial bodies. Chaos and Correlation, November 1st, 2006.
17. Татьяна Черноглазова, Александр Трунев. ХРОНИКИ АТОМНЫХ КАТАСТРОФ 20 ВЕКА/ Chaos and Correlation, March 11, 2008. <http://trunev.com/Chaos/March2008/AS/AS.htm>