

УДК 303.732.4

UDC 303.732.4

**СИСТЕМА ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА ДЛЯ ИДЕНТИФИКАЦИИ
СОЦИАЛЬНЫХ КАТЕГОРИЙ
РЕСПОНДЕНТОВ НА ОСНОВЕ
АСТРОНОМИЧЕСКИХ ПАРАМЕТРОВ**

**ARTIFICIAL INTELLIGENCE SYSTEM FOR
IDENTIFICATION OF SOCIAL CATEGORIES
OF NATIVES BASED ON ASTRONOMICAL
PARAMETERS**

Евгений Луценко
д.э.н., к.т.н., профессор

Eugene Lutsenko
Dr. Sci. Econ., Cand. Tech. Sci., Prof.

*Кубанский государственный аграрный
университет, Краснодар, Россия*

*Kuban State Agricultural University, Krasnodar,
Russia*

Александр Трунев
д. ф.-м.н.

Alexander Trunev
Dr. Sci. Phys.-Math., Ph.D

*Директор, A&E Trounev IT Consulting, Торонто,
Канада*

*Director, A&E Trounev IT Consulting, Toronto,
Canada*

В статье дается обзор когнитивного моделирования записей AstroDatabank с использованием системы искусственного интеллекта «Эйдос». Обсуждаются наиболее важные результаты и технология моделирования.

The cognitive simulation of AstroDatabank records by using the Artificial Intelligence System – AIDOS is reviewed in this paper. The technology of simulation is described and the mostly important results are discussed.

Ключевые слова: СЕМАНТИЧЕСКИЕ
ИНФОРМАЦИОННЫЕ МОДЕЛИ,
АСТРОДАТАБАНК, АСТРОНОМИЧЕСКИЕ И
СОЦИОЛОГИЧЕСКИЕ БАЗЫ ДАННЫХ,
ОБУЧЕНИЕ НЕЙРОСЕТЕЙ, ЦИФРОВОЙ
ЭКСПЕРИМЕНТ.

Keywords: SEMANTIC INFORMATION MODELS,
ASTRODATABANK, ASTRONOMICAL AND
SOCIOLOGICAL DATABASES, NEURON-NET
TRAINING, NUMERICAL EXPERIMENT.

Introduction

New method of identification of a birth chart based on system-cognitive analysis and on the advanced information theory [1] was developed recently [2–3]. This method differs from the normal astrological models so that the birth chart is not interpreted, but it is identified by using a number of attributes and categories, by comparing with the astrological database [4–5], which includes a description of the many key events in real life of real persons. As a result of the identification each person receives a customized description contains classes and categories of events, indicating the likelihood of their implementation. In this research not used any astrological interpretation or any astrological rules. Statistical patterns and the correlation revealed in the data processing of the artificial intelligence system by comparing birth charts and biography. Test examples demonstrate the effectiveness of the system for the recognition of certain classes of entities.

Input Databases

The main source of astrological database prepared for the artificial intelligence system simulation is the original (first version) Lois Rodden's AstroDatabank [4] and AstroDatabank v. 4.0 [5]. These databases contain biography of famous and ordinary people so that all the categories and events of life are classified and ordered.

Data imported from AstroDatabank v. 4.0 were converted into a DBF4 format database. Only 9897 records have been utilized including 5 categories shown below with corresponding number of records:

Table 1: Four classes, 5 categories and related number of records

KOD_OBJ	NAME	ABS
1	Politics, Science	1876
2	Medical: Physician	347
3	Sports	6032
4	Psychological	1642

Note, 184 records are repeated among 9897 since they related to 2, 3 or 4 categories listed above. Records were cooperated in four classes as shown in Table 1. Every record has 23 active numerical cells consist of coordinates of celestial bodies, Ascendant and Midhaven at the moment of birth and in the place of birth, i.e.:

- Longitude (degree) of the Sun, the Moon, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto, North Node, Ascendant and Midhaven;
- Declination (degree) of the Sun, the Moon, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto.

From this database were derived two databases to study a declination effect on the similarity parameter:

1. Database1 with 23 active numerical cells in every of 9897 records as described above but all Declination parameters were adapted to the longitude interval (0; 360) by using formula: $\text{Declination1} = (\text{Declination} + 30) * 6$.
2. Database0 with 23 active numerical cells in every of 9897 records as described above, but all Declination parameters were recalculated as follows: $\text{Declination0} = \text{Declination} * 0$, also for all records we put $\text{Ascendant} = \text{Midhaven} = 0$, therefore only Longitude of the Sun, the Moon, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto and North Node have been utilized in this database.

After this minor adaptation all 23 cells have one scale and format, therefore they could be analyzed in the same manner as well as the declination parameter effect on the simulated outcomes could be studied.

The data imported from original Lois Rodden's AstroDatabank were converted into the Borland JDataStore format databases. Then, the data were sorted using SQL queries and special functions written in Java. Only 20007 records related to 1931 categories and events have been utilized in this research. For these records were calculated coordinates of celestial bodies (latitude and longitude in degrees, and the distance in astronomical units). 12 cusps of astrological houses in the Placidus system were calculated for records with the exact time of birth. The ephemerides following celestial bodies and points were established: the Sun, the Moon, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto, and North Node. The next step is sorting by category of records. As result XML tree categories reference database was obtained. Next, the database has been completely exported in Excel and then it converted to the DBF4 format (which accepted by the artificial intelligence system). Only 23 active numerical cells in every of 20007 records were utilized in this research, i.e.: Longitude (degree) of the Sun, the Moon, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto, North Node, and 12 cusps of the astrological houses (houses in the Placidus system). From this database were derived several databases:

1. Database A of 20007 records related to 500 representative categories (category represented in the database at least 26 times).
2. Database B of 15007 records related to 500 representative categories – training data set.
3. Database C of 5000 records which are not used in the Database B (but used in the Database A) – recognized data set.
4. Database D of 20007 records related to 240 unrepresentative categories (number of records related to category higher than 2 and less than 25) – low frequency limit.
5. Database E of 20007 records related to 870 categories (number of records related to any category higher than 2) – mostly complete database.
6. Database F of 20007 records related to 37 categories (number of records related to any category higher than 1000) – higher frequency limit.
7. Database F1 of 20007 records related to 100 categories (number of records related to any category higher than 174).

8. Database G of 20007 records related to 4 categories listed below in Table 2, b. In this database 8150 records are not involved in a simulation.

Table 2: Four classes, four categories and related number of records in a case of Database G

KOD_OBJ	NAME	ABS
1	Famous	3373
2	Medical	2910
3	Sports	4567
4	Psychological	1007

Note 20007 records are related to the original (first version) Lois Rodden's AstroDatabank [4] and AstroDatabank v. 4.0 [5] as well. The difference between these databases is that latest version updated with more than 5000 records, and it is a reason why the same category SPORT has different records in Table 1 and 2.

The Model and the Artificial Intelligence System - AIDOS

As well know there are several ways to decompose Zodiac circle in a process of analyzing a birth chart:

- day and night houses partition – 2 sectors;
- Cardinal signs, fixed and mutable signs – 3 multiply connected sectors.
- squares – 4 sectors;
- partition based on element of fire, earth, air and water – 4x3 sectors;
- zodiac signs – 12 sectors;
- decants – 36 sectors;
- terms – 60 sectors;
- Degree – 360 sectors.

Decomposition combinations such as those listed above seem to resemble algorithms of grid simulation widely used in a modern science, in which condensation of the grid helps improve convergence in solving the task. We utilized this method in order to perform packet recognition of 9897 or 20,007 records exported from AstroDatabank and presented as DBF4 format databases. In order to do this a solution founded based on data from 172 grids of various dimensions, containing 2, 3, 4, .., 173 sectors consequently (it is a limit for this task at the moment). Thus the net entropy effect could be established during this simulation with the system of artificial intelligence AIDOS [2].

Standard AIDOS package includes 7 subsystems and 85 programmable applications organized in a block structure – see Table 3. Generally speaking it is a neuron-net computer application running under Windows XP in MS-DOS mode, designed with CLIPPER 5.01, Tools-II and BiGraph 3.01, provided the following objectives:

1. Synthesis and adaptation of the semantic data model.
2. Identification and forecasting.
3. Precise analysis of the semantic data model.

Table 3: Generalized Structure of the Universal Cognitive Analytical System AIDOS, v. 12.03.2008

Subsystem	Mode	Function	Operation
1. Formalization	1. The classification scale and graduation		
	2. Descriptive scale (and graduation)		
	3. Graduation descriptive scales (signs)		
	4. Hierarchical systems levels	1. Levels of classes	
		2. Levels of signs	
	5. Software interfaces for importing data	1. Import data from TXT files-standard DOS-text	
		2. Import data from DBF files (Standard Prof. A. N. Lebedev)	
		3. Imports from transposed DBF files (Standard Prof. A. N. Lebedev)	
		4. Generation scales and training set RND model	
		5. Generation scales and training sample for the numerical study	
		6. DBF-matrix transposition of baseline data	
		7. Import data from DBF files (Standard E. Lebedev)	
	6. Postal Service to INS	1. Exchange grade	
2. Exchange of generalized signs			
3. Exchange of primary signs			
7. Printing questionnaire			
2. Synthesis SIM	1. Writing-adjustment training set		
	2. Management of training sample	1. Parametric objects for processing job	
		2. Statistical parameters, hand sample	
		3. Auto sample of training set	
	3. Synthesis of semantic data model SIM	1. The calculation of the absolute frequency matrix	
		2. Excluding artifacts (robust procedure)	
		3. Calculation information matrix SIM-1 and converting into executable information matrix	
		4. The calculation of conditional interest distributions SIM-1 and SIM-2	
		5. Automatic execution regimes 1-2-3-4	
		6. Measurement of convergence and stability model	1. Convergence and sustainability SIM
2. Dependence validity of the model training set			
7. Calculation information matrix SIM-2 and converting into executable information matrix			
4. Postal Service to educational information			
3. Optimizing SIM	1. Formation of the classes orthonormal basis		
	2. Excluding signs of a low selective force		
	3. Removing classes and attributes for which insufficient data		
	4. Divisions on the part of the typical and atypical		
	5. Generation of associated signs and convert a training sample		
4. Transcribing	1. Writing-adjustment recognizable sample		
	2. Batch recognition		
	3. The withdrawal of recognition results	1. Cut: an object - a lot of classes	
2. Cut: one class - many sites			

		4. Postal Service recognizable sample	
		5. Construction of the functions of influence	
		6. Decoding combinations recognizable signs in the sample	
5. Typology	1. Typological analysis classes recognition	2. Classes cluster and constructive analysis	1. Information (rank) portraits (classes)
			1. The calculation of similarity matrix of classes images
			2. Generation of clusters and constructs classes
			3. Viewing and printing cluster, and constructs
			4. Automatic execution modes: 1,2,3
			5. Conclusion 2D semantic networks classes
			3. Cognitive charts of classes
	2. Typological analysis of the primary signs	2. Signs cluster and constructive analysis	1. Information (rank) portraits of signs
			1. The calculation of similarity matrix of signs images
			2. Generation of clusters and constructs signs
3. Viewing and printing cluster, and constructs			
4. Automatic execution modes: 1,2,3			
		5. Conclusion 2d semantic networks signs	
		3. Cognitive signs of charts	
6. Semantic-Cognitive analysis of model SIM		1. Estimation of the objects completion reliability	
		2. The measurement of the adequacy of semantic data model	
		3. Measuring independence classes and signs	
		4. Viewing profiles classes and signs	
		5. Graphic display of non-local neurons	
		6. Displaying subsets of the neural network	
		7. Classical and Integral cognitive maps	
7. Service	1. Generation Databases (dumping)	2. NSI	1. All database
			1. All databases NSI
			2. Classes DB
			3. Initial signs DB
			4. Generalized signs DB
			3. Training sample
			4. Recognized sample
			5. Statistics Database
			2. Reload index all databases
			3. Print database absolute frequencies
			4. Printing conditional interest distributions Databases SIM-1 and SIM-2
			5. Printing information of SIM-1 and SIM-2 Databases
			6. Descriptive Intelligent Information Retrieval System
			7. Copying the major SIM databases
			8. Convert SIM-1 into executable information matrix
		9. Convert SIM-2 into executable information matrix	

The cognitive simulation of AstroDatabank records including the neuron-net training and recognition was realized for any grid of fixed dimension $N=2, 3, 4, \dots, 173$ sectors. Thus there are many models – M2, M3, M4... M173 corresponding to the number of sectors in a given partition of Zodiac. For every model could be established own catalog (they are numbered simply as 002, 003, 004 ...) and a copy of the system AIDOS. To manage the input parameters and outcomes of all models a special system has been designed [3], which would be implemented "collectives decisive rules» i. e., would the ability to automatically generate a number of models that would form one coherent system, which called "multi-model". This system consists of few programmable applications which allow setup any combination of models; run the neuron-net training and

recognition for all models, organize and summarize the results of the identification of the respondents in different models for a set of categories.

Main Results

The technology of simulation described in papers [6–8]. In fact the system AIDOS operates with Object Code like numbers in a left column in Tables 1, 2. Astronomical parameters also have own code called “scale or graduation code”, for instance, in a case of model M3 we have 23 main scales and 69=23*3 graduations; six of them shown below:

Code	Name of scale
1	SUNLON-[3]: {0.000, 120.000}
2	SUNLON-[3]: {120.000, 240.000}
3	SUNLON-[3]: {240.000, 360.000}
4	MOONLON-[3]: {0.000, 120.000}
5	MOONLON-[3]: {120.000, 240.000}
6	MOONLON-[3]: {240.000, 360.000}

If any record in a training database shows a longitude of the Sun belongs to the interval (0.000, 120.000) then a frequency of the corresponding code 1 increases on a unit. Therefore a frequency of scales in the training database could be calculated and the frequency matrix and the information matrix could be established. For example, in a case of model M2 trained with Database F, a fragment of the frequency matrix and a fragment of the information matrix are shown in Table 4 and 5 consequently:

Table 4: The frequency matrix (fragment) in a case of model M2 trained with Database F (frequency is given in absolute value) [7]

Code of scale	Code of category											
	1	2	3	4	5	6	7	8	9	10	11	12
1	6744	2623	2281	2201	1671	1477	1378	1271	1222	1201	1230	1208
2	6896	2502	2286	2270	1702	1433	1297	1306	1220	1195	1155	1152
3	6786	2539	2325	2187	1689	1445	1330	1273	1207	1211	1218	1177
4	6854	2586	2242	2284	1684	1465	1345	1304	1235	1185	1167	1183
5	6261	2401	2070	2039	1561	1343	1307	1185	1125	1086	1134	1156
6	7379	2724	2497	2432	1812	1567	1368	1392	1317	1310	1251	1204
7	6907	2688	2332	2274	1735	1510	1422	1301	1263	1193	1232	1263
8	6733	2437	2235	2197	1638	1400	1253	1276	1179	1203	1153	1097
9	7137	2760	2443	2344	1754	1500	1454	1341	1269	1223	1330	1279
10	6503	2365	2124	2127	1619	1410	1221	1236	1173	1173	1055	1081

Actually an information counted in the system with 8 decimal places, but in Table 5 it shown with 2 decimal position (*100) only. A positive or negative value of information in a cell ij in Table 5 means that category j has a positive or negative correlation with scale i.

Table 5: The information matrix (fragment) in a case of model M2 trained with Database F (information given in Bit*100) [7]

Code of scale	C o d e o f c a t e g o r y											
	1	2	3	4	5	6	7	8	9	10	11	12
1	3	-1	17	-2	-3	-0	-3	-3	-3	-3	22	-4
2	4	-3	17	-0	-3	-1	-5	-2	-3	-3	20	-5
3	3	-2	18	-2	-3	-1	-4	-3	-3	-3	22	-4
4	3	-2	16	-0	-3	-1	-4	-2	-2	-4	20	-5
5	3	-2	16	-2	-3	-1	-2	-3	-3	-4	22	-2
6	3	-3	17	-1	-3	-1	-6	-2	-2	-3	20	-6
7	3	-1	17	-1	-3	-1	-3	-3	-2	-4	21	-3
8	4	-3	17	-1	-3	-1	-5	-2	-3	-2	21	-6
9	3	-1	17	-1	-3	-2	-3	-3	-3	-4	23	-3
10	3	-3	16	-1	-3	-0	-5	-2	-2	-2	19	-6

When a training of the neuron-net for every model is finished, then packet recognition could be run. It starts from definition of recognized sample records number. In a case of Database0, Database1 or Database G with 4 classes only a reasonable number could be N=400 or 100 per class. The trained computer neuron-net has a reaction on any input data which are similar to the training sample. Therefore every record from N could be analyzed and four possible reactions on it could be measured:

- Record with number **n** from N belongs to the category number **m** and it's true, the correlation parameter of the record number **n** with the category number **m** is BT_{nm} ;
- Record with number **n** from N not belongs to the category number **m** and it's true, the correlation parameter is T_{nm} ;
- Record with number **n** from N belongs to the category number **m** and it's false, the correlation parameter is BF_{nm} ;
- Record with number **n** from N not belongs to the category number **m** and it's false, the correlation parameter is F_{nm} ;

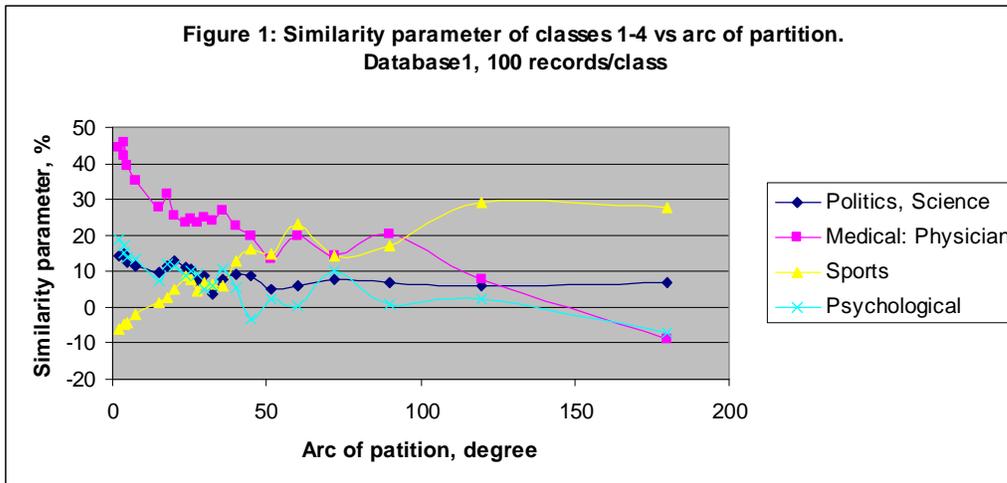
Thus the effective artificial intelligence system should be designed in a way to minimize a false prediction and to maximize a true prediction. For the best understanding of the packet recognition results a special form of the similarity parameter has been proposed as follows [7]:

$$S_m = \frac{1}{N} \sum_{n=1}^N (BT_{nm} + T_{nm} - BF_{nm} - F_{nm}) \cdot 100\% .$$

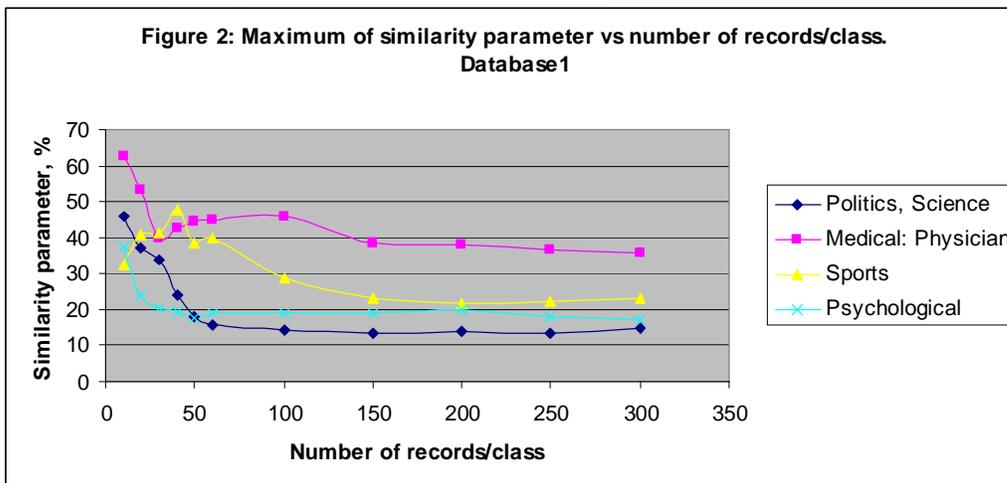
With this definition the similarity parameter changes from -100 % up to 100 %, like a statistical correlation parameter. If $S_m=0$, it means that the category number **m** is not recognized well even if $BT_{nm} = 0.95$ for every true record and it looks like a very good result. From the other side if $S_m=0.5$, it is really a good result even if $BT_{nm} = 0.5$ for every true record, but it means that there are no false records and every true record been recognized. Let conduct few experiments to recognize several categories.

EXPERIMENT 1

In the first experiment the multi-model of 22 models including M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15, M18, M20, M24, M48, M72, M90, M96, M150 was setup and then 22 models were trained with Database1 of 9897 records. As result an information image (portrait) of every class has been simulated. Similarity parameters of classes 1–4 (series 1–4) from Table 1 versus the arc of partition (degree) in a case of packet recognition 100 records/class are shown in Figure 1. The number of records effect on the similarity parameter shown in Figure 2, where data for the maximum of the similarity parameter are plotted.



Class	NAME	ABS
1	Politics, Science	1876
2	Medical: Physician	347
3	Sports	6032
4	Psychological	1642

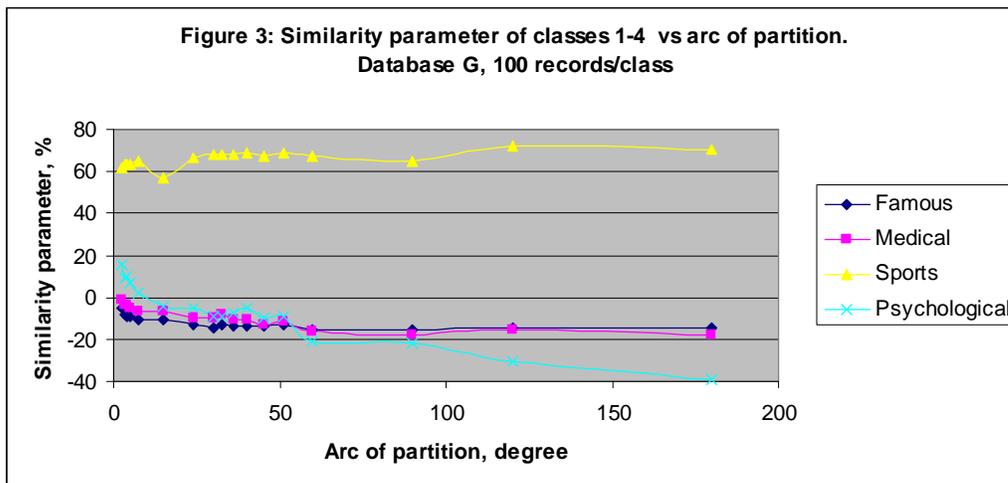


In the first experiment the best result obtained for the category “Medical: Physician” - $S = 45.908\%$ in a case of model M90 and for 100 records/class. Reducing a number of records/class it is possible to increase a similarity parameter of the category “Medical: Physician” up to 62.722 in a case of model

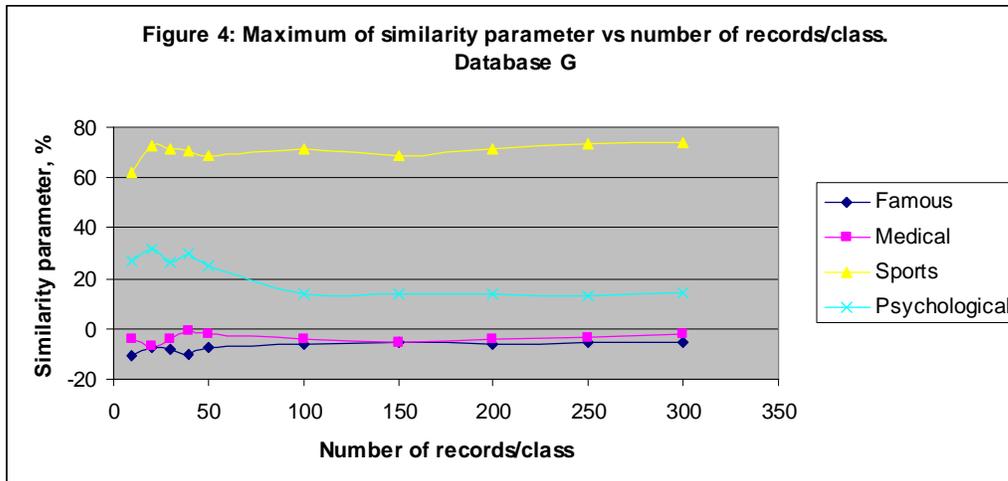
M150 and for 10 records/class – see Figure 2. For the category “Sport” the best result $S= 47.526 \%$ was found in a case of model M4 for 40 records/class. Note that it is less than a random choice probability = 0.609478 for this category. Nevertheless, a similarity parameter reflects a response of the artificial intelligence system on the astronomical parameters effect on the training and recognition while a random choice probability is a fixed value for a fixed database, and it depends on the number of records only.

EXPERIMENT 2

In the second experiment all simulations of the first experiment have been repeated with Database G of 20007 records – see Figures 3–4. In this experiment the best recognized category is “Sport” with $S= 72.273$ in a case of model M3 and for 100 records/class.

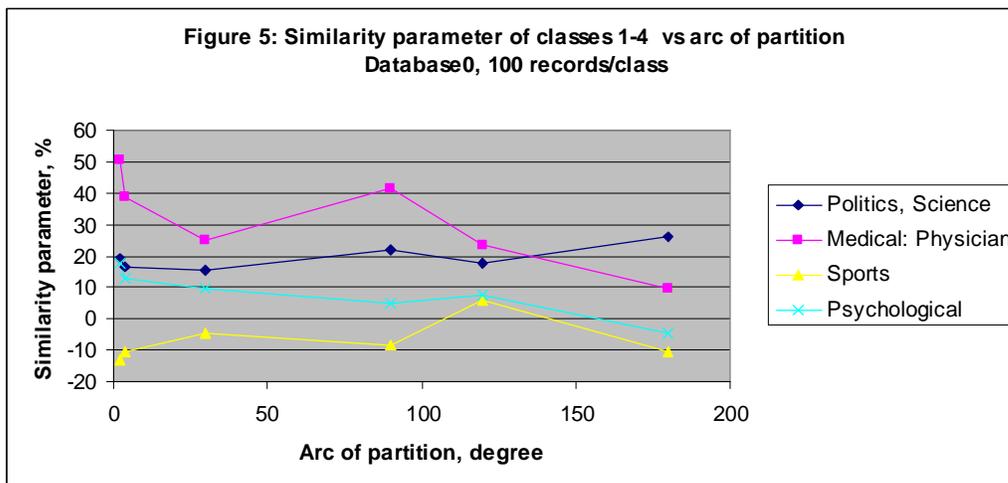


Class	NAME	ABS
1	Famous	3373
2	Medical	2910
3	Sports	4567
4	Psychological	1007



EXPERIMENT 3

In the third experiment the multi-model of 6 models including M2, M3, M4, M12, M90, and M150 was established and trained with Database0 (9897 records). Similarity parameters of classes 1–4 (series 1–4) from Table 1 versus the arc of partition (degree) in a case of packet recognition 100 records/class are shown in Figure 5. There is a big difference in the final results for two databases – Database1 and Database0 (see Figure 1 and Figure 5); even they have identical number of records, but different number of scales – 23 (longitude and declination of 10 planets, longitude of North Node, Ascendant and MC) and 11 (longitude of 10 planets and North Node only) consequently. In this experiment the best result was found for the category “Medical: Physician” - $S=50.634\%$ in a case of model M150, and it is comparable with data shown in Figure 1. For the category “Sport” the best result is $S=5.915\%$ in a case of model M3, and it is much less than $S=28.935\%$ found for this category in a case of Database1 and model M3 – see Figure 1.



Class	NAME	ABS
1	Politics, Science	1876

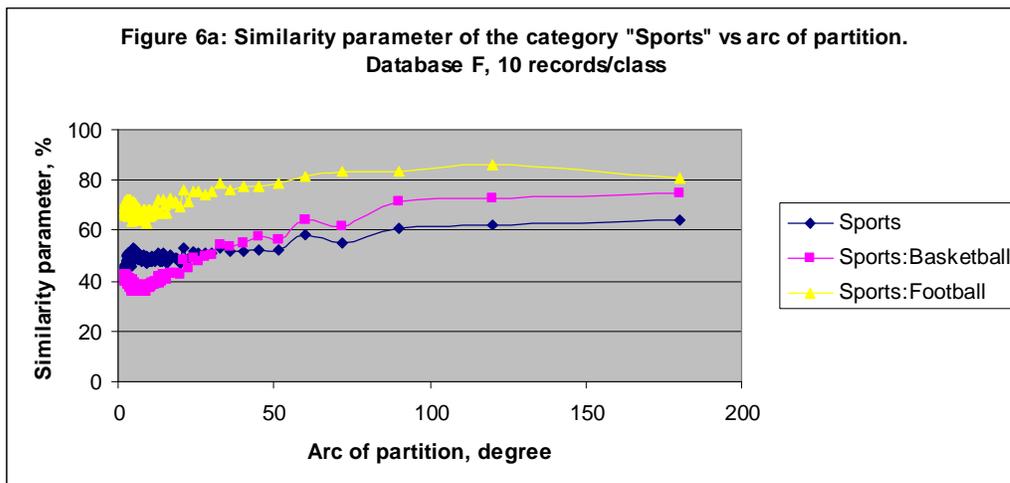
2	Medical: Physician	347
3	Sports	6032
4	Psychological	1642

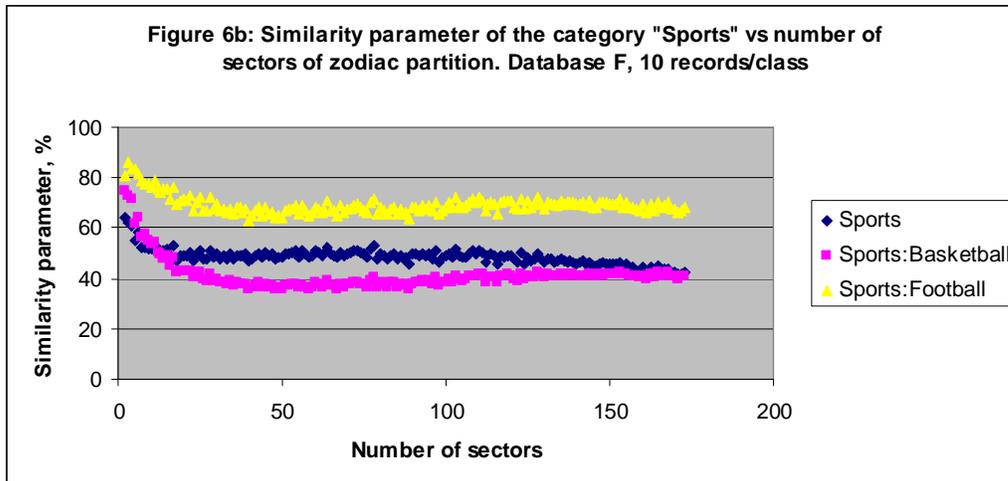
EXPERIMENT 4

In the fourth experiment the multi-model of 172 models including M2, M3, M4, ..., M172, and M173 was established and trained with Database F (20007 records) [7]. With this model it is possible to run a precise simulation for those categories which been decomposed in several subcategories or classes. For instance, the similarity parameter of the category “Sports” decomposed in three classes (see Table 6) shown in Figures 6a, 6b versus arc of partition and number of sectors of zodiac circle partition consequently. The best result $S= 85.864$ found for the subcategory “Sports: Football” in a case of model M3.

Table 6: The category “Sports” decomposed in three classes and related numbers of records. Database F

Class	NAME	ABS
1	Sports	4567
2	Sports: Football	1613
3	Sports: Basketball	2385



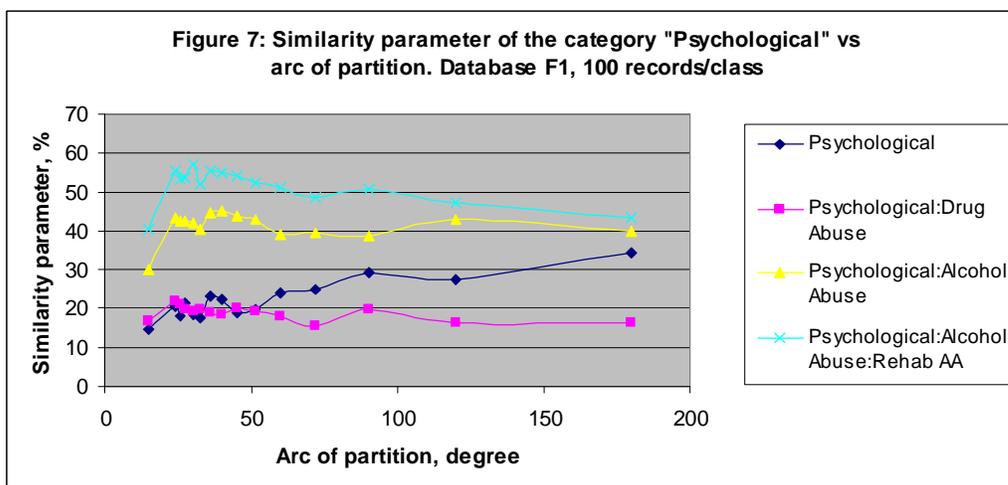


EXPERIMENT 5

In this experiment the multi-model of 15 models including M2,M3,M4,M5,M6,M7, M8,M9,M10,M11,M12,M13,M14,M15,M24 was established and trained with Database F1 (20007 records). The similarity parameter of the category “Psychological” decomposed in four classes – see Table 7, shown in Figure 7. The best result S= 57.244 found for the subcategory “Psychological: Alcohol Abuse: Rehab AA” in a case of model M12. Note that subcategories mostly showed better results in recognition than a main category.

Table 7: The category “Psychological” decomposed in four classes and related numbers of records. Database F1

Class	NAME	ABS
1	Psychological	1007
2	Psychological:Drug Abuse	282
3	Psychological:Alcohol Abuse	481
4	Psychological:Alcohol Abuse:Rehab AA	267

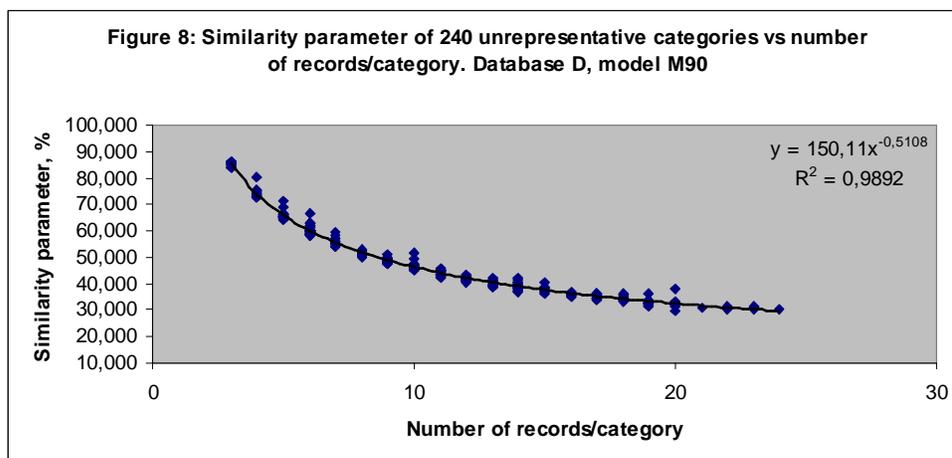


EXPERIMENT 6

In this experiment the model M12 only has been setup and trained with Database B of 15007 records. Then all records from Database C have been utilized for recognition. In result the number of the true recognized records was determined as $N_{true}=3435$ or 68.7 % of 5000 records [6]. To compare this result with some background data the stochastic Database of 5000 records has been generated (all the active sell numbers taken from a random set) the same size and format as Database C, and recognized, finally a maximum of the similarity parameter has been established as $S_{max}=1.206$ % [6]. Therefore a value of the similarity parameter which is higher than $1.206 * 2.5 = 3.015$ % should be considered as a certain value with 95 % probability. This criterion was taken into account in the simulation with records of Database C.

EXPERIMENT 7

In seven experiment the multi-model of 4 models including M3, M4, M12 and M90 was setup and trained with Database D of 20007 records related to 240 unrepresentative categories (number of records related to category higher than 2 and less than 25). The similarity parameter of 240 categories versus the number of records related to every category in Database D in a case of the model M90 shown (together with a trend line) in Figure 8. These data illustrate the low frequency trend in a case of the recognition, when the number of records for any category is not statistically representative.



EXPERIMENT 8

In this experiment the multi-model of 16 models including M2,M3,M4,M5,M6,M7, M8,M9,M10,M11,M12,M24,M36,M48,M60 and M72 was established and trained with Database E (20007 records and 870 categories).

The similarity parameter of 870 categories versus the number of records related to every category in Database E in a case of the model M72 shown (together with a trend line) in Figure 9 (there is a double logarithmic scale performed). The trend line in this case has the same slope like in Figure 8 therefore it could be a common correlation for 20007 records utilized in both databases - D and E. The similarity parameter of the category “Medical” and several subcategories (see Table 8) are shown in Figure 10. The best result $S=65.109$ found for the subcategory “Medical: Doctor: Therapist” in a case of model M3.

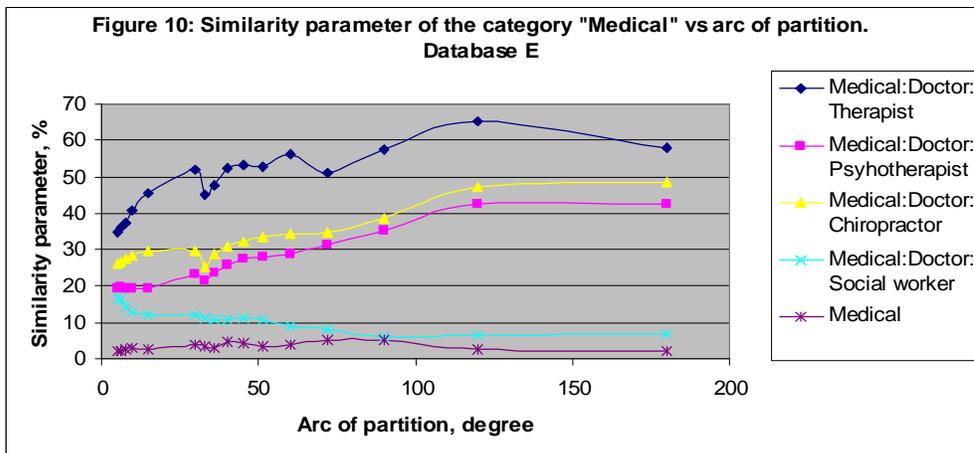
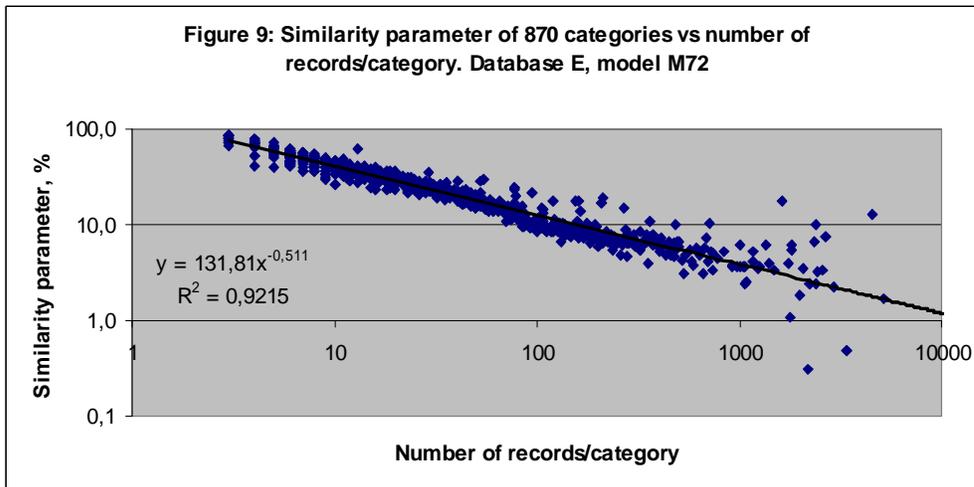


Table 8: Category “Medical”, subcategories and related number of records

Medical:Doctor:Therapist	29
Medical:Doctor:Psychotherapist	79
Medical:Doctor:Chiropractor	33
Medical:Doctor:Social worker	54
Medical	2910

EXPERIMENT 9

In this experiment a multi-model of 10 models including M3,M4,M5,M6,M9, M12,M15,M18,M20,M24 was trained with Database A of 500 representative categories (category represented in the database at least 26 times). The similarity parameter of the category “Death: Long life >80 yrs”, and several subcategories (see Table 9) are shown in Figure 10. The best result $S=27.504$ found for the subcategory “Age 89” in a case of model M4. Note that all subcategories data shown in Figure 11 have synchronic behavior versus the arc of partition.

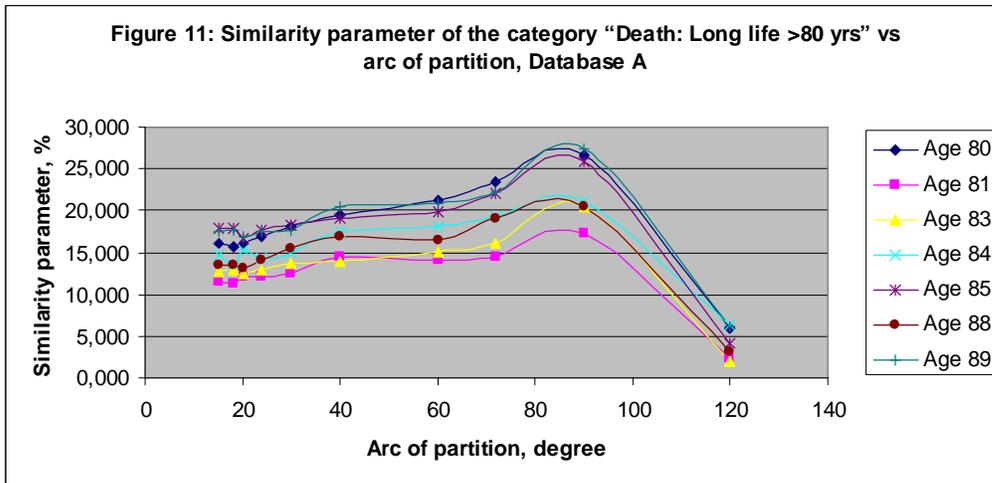
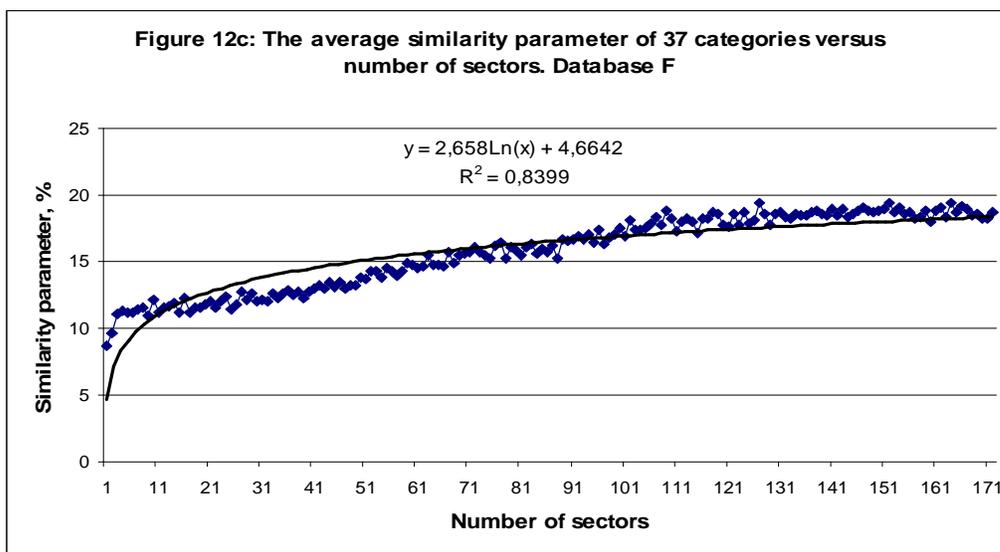
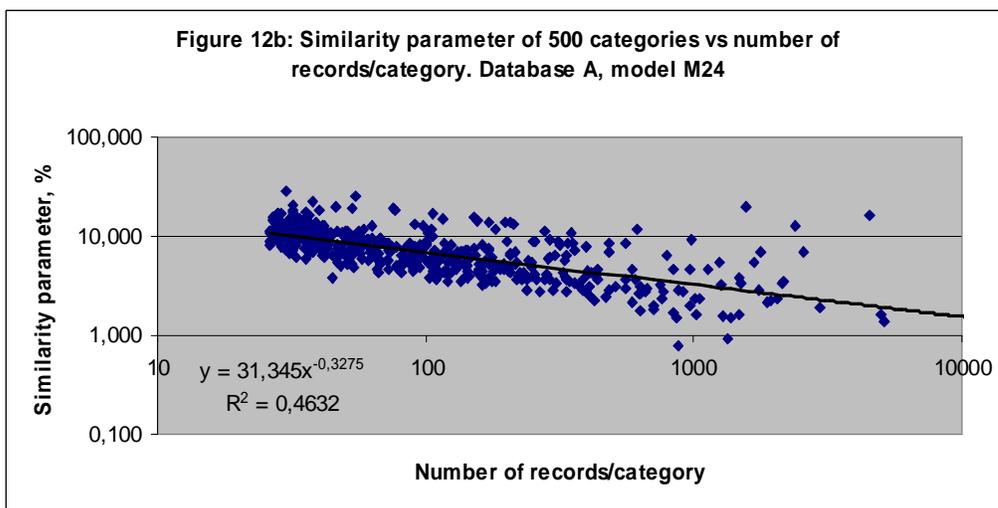
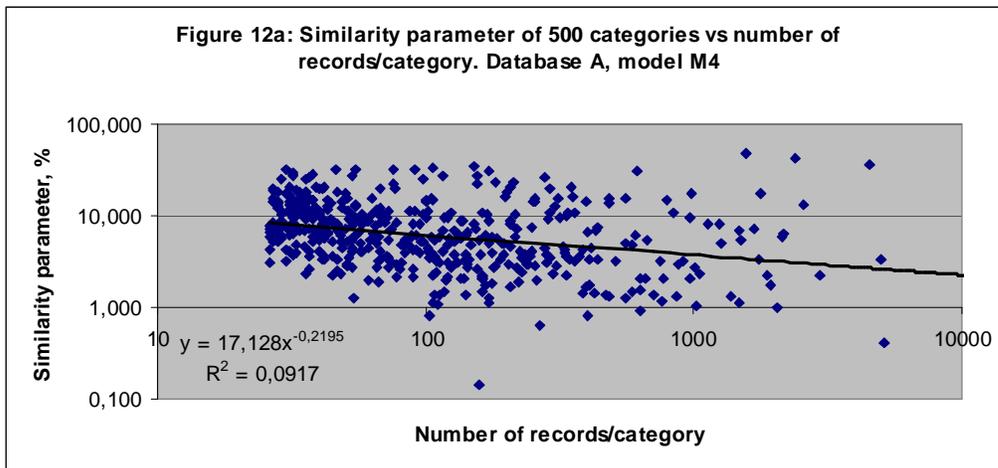


Table 9: Category “Death: Long life >80 yrs”, subcategories and related number of records

Age 80	37
Age 81	50
Age 83	42
Age 84	31
Age 85	36
Age 88	39
Age 89	32

Net entropy effect on the similarity parameter

The similarity parameter of 500 categories versus the number of records related to every category in Database A in a case of the model M4 shown (together with a trend line) in Figure 12a. These data look like chaotically dispersed points. There is a dramatic difference between data in Figures 9 and Figure 12. It should be noted that both databases A and E have same numbers of records per category but there are different numbers of scales which depend on number of sectors. Therefore it is possible to increase a correlation by increasing number of scales see Figure 12b. It calls the net entropy effect.



In Figure 12c the average similarity parameter of 37 categories versus number of sectors is shown [7]. These data could be approximated by the logarithmic function – a solid line in Figure 12c. The function of entropy (or information) also depends on the number of elements as a logarithmic function [1]. Thus the

average similarity parameter is a linear function of the net entropy (or information as well). Nevertheless some categories better recognized at the small number of sectors – see Figure 6b for instance.

Discussion

Several databases have been tested with the artificial intelligence system AIDOS to found out the astronomical parameters effect on the social categories of natives. The data of the multi-model simulations shown in Figures 1–10 demonstrate a regular respond of the similarity parameter on variations of the number of records per class or category as well as on the arc of zodiac cycle partition. Therefore the astronomical parameters effect on the social categories of natives could be investigated and determined as it has been performed above.

An information portrait is the main astronomical characteristic of any category. For instance, the category “Sports” in a case of Database F1 and model M12 could be characterized as follows (only 62 scales of 276 are shown):

Scales	Information, Bit
PLUTOLON-[12]: {150.000, 180.000}.	0.832
URANUSLON-[12]: {150.000, 180.000}.	0.815
NEPTUNELON-[12]: {210.000, 240.000}.	0.808
URANUSLON-[12]: {120.000, 150.000}.	0.529
SATURNLON-[12]: {330.000, 360.000}.	0.480
SATURNLON-[12]: {0.000, 30.000}.	0.448
URANUSLON-[12]: {180.000, 210.000}.	0.433
NODELON-[12]: {0.000, 30.000}.	0.371
SATURNLON-[12]: {300.000, 330.000}.	0.362
NODELON-[12]: {30.000, 60.000}.	0.338
JUPITERLON-[12]: {120.000, 150.000}.	0.324
NODELON-[12]: {90.000, 120.000}.	0.324
NODELON-[12]: {60.000, 90.000}.	0.323
SATURNLON-[12]: {270.000, 300.000}.	0.304
MARSLON-[12]: {150.000, 180.000}.	0.299
JUPITERLON-[12]: {150.000, 180.000}.	0.298
NODELON-[12]: {120.000, 150.000}.	0.291
MERCURYLON-[12]: {180.000, 210.000}.	0.288
MOONLON-[12]: {120.000, 150.000}.	0.279
MARSLON-[12]: {90.000, 120.000}.	0.279
VENUSLON-[12]: {210.000, 240.000}.	0.277
JUPITERLON-[12]: {180.000, 210.000}.	0.272
JUPITERLON-[12]: {30.000, 60.000}.	0.272
SUNLON-[12]: {210.000, 240.000}.	0.270
JUPITERLON-[12]: {60.000, 90.000}.	0.261
MARSLON-[12]: {120.000, 150.000}.	0.260
SUNLON-[12]: {150.000, 180.000}.	0.255
VENUSLON-[12]: {0.000, 30.000}.	0.253
MOONLON-[12]: {270.000, 300.000}.	0.253
SUNLON-[12]: {120.000, 150.000}.	0.247
JUPITERLON-[12]: {0.000, 30.000}.	0.247
MERCURYLON-[12]: {120.000, 150.000}.	0.246

VENUSLON-[12]: {180.000, 210.000}.	0.246
VENUSLON-[12]: {60.000, 90.000}.	0.245
SUNLON-[12]: {0.000, 30.000}.	0.245
MERCURYLON-[12]: {60.000, 90.000}.	0.244
JUPITERLON-[12]: {90.000, 120.000}.	0.244
MARSLON-[12]: {240.000, 270.000}.	0.243
VENUSLON-[12]: {120.000, 150.000}.	0.241
MOONLON-[12]: {0.000, 30.000}.	0.238
MERCURYLON-[12]: {330.000, 360.000}.	0.238
MOONLON-[12]: {60.000, 90.000}.	0.238
VENUSLON-[12]: {150.000, 180.000}.	0.233
MERCURYLON-[12]: {300.000, 330.000}.	0.232
MERCURYLON-[12]: {150.000, 180.000}.	0.231
SUNLON-[12]: {330.000, 360.000}.	0.231
MOONLON-[12]: {150.000, 180.000}.	0.224
NEPTUNELON-[12]: {180.000, 210.000}.	0.222
MARSLON-[12]: {180.000, 210.000}.	0.221
NODELON-[12]: {150.000, 180.000}.	0.221
MOONLON-[12]: {90.000, 120.000}.	0.221
MOONLON-[12]: {330.000, 360.000}.	0.218
MOONLON-[12]: {240.000, 270.000}.	0.218
VENUSLON-[12]: {330.000, 360.000}.	0.217
SUNLON-[12]: {240.000, 270.000}.	0.216
MERCURYLON-[12]: {240.000, 270.000}.	0.215
SUNLON-[12]: {180.000, 210.000}.	0.214
MARSLON-[12]: {330.000, 360.000}.	0.214
MERCURYLON-[12]: {270.000, 300.000}.	0.213
VENUSLON-[12]: {300.000, 330.000}.	0.211
SUNLON-[12]: {90.000, 120.000}.	0.210
MARSLON-[12]: {210.000, 240.000}.	0.209

It is impossible to derive any simple suggestion like “category Sports depends on the Pluto or Mars position” from this portrait only. Generally speaking any information portrait depends on the utilized model and database. Nevertheless it gives some ideas about predominate scales in the information portrait of the category Sport. Note that every scale contributes a portion of information which actually utilized for recognition.

Every recognized record has own similarity portrait, for example, the portrait of the record for Bush, George Walker could be presented as follows:

Bush, George Walker	
Category	Similarity parameter, %
SC: B795-Sports: Bullfighting.	63
SC: C1070-Travel: Crew/ Ship, Train, Bus: Taxi driver.	63
SC: C619-Education: Teacher: Language/English.	56
SC: C604-Work: Food and Beverage: Alcohol business.	47
SC: C1340-Medical: Doctor: Chiropractor.	46
SC: D1256-Education: Teacher: Science: Computer science.	45

SC:C382-Medical:Doctor:Dentist/ Dental Tech.	45
SC:B236-Business:CPA/ Auditor/ Accountant.	42
SC:B1023-Law:Court reporter.	37
SC:C421-Occult Fields: Psychic/ Medium/ Spiritualist: Palmist.	37
SC:C614-Medical:Doctor:Veterinarian.	36
SC:C603-Work:Food and Beverage: Farmer/ Rancher.	36
SC:D1529-Business:Business/Marketing:Real estate: Agent.	31
SC:C901-Work:Food and Beverage: Fast-food service.	31
SC:C620-Business:Entertain/Business:Manager/ Agent.	30
SC:D1246-Education:Teacher:Science:Philosophy.	30
SC:C887-Work:Maintenance Field: Clerk.	29
SC:C461-Work:Food and Beverage: Chef/ Cook.	29
SC:C570-Travel:Adventurer:Explorer.	29
SC:C62-Entertainment:Music:Group/ Duo.	28
SC:B169-Medical:Doctor.	27
SC:C422-Occult Fields: Psychic/ Medium/ Spiritualist: Tarot reader.	27
SC:C551-Famous:Greatest hits: Science field.	26
SC:A99-Financial.	26
SC:C575-Work:Maintenance Field: Factory work.	26
SC:C782-Science:Biology:Zoology.	26
SC:C1257-Education:Teacher:Coach.	24
SC:C267-Entertainment:Music:Song writer.	24
SC:C1038-Education:Teacher:Science.	24
SC:D791-Medical:Doctor:Psychologist:Parapsychology.	24
SC:B628-Science:History.	24
SC:C1002-Military:Military service.	23
SC:B32-Business:Business owner.	23
SC:D104-Entertainment:Music:Vocalist:Opera.	22
SC:C263-Politics:Heads of state: U.S. Presidents.	22
SC:B413-Entertainment:Child performer.	22
SC:B626-Occult Fields: Out of Body experience.	21
SC:C375-Business:Sports Business: Coach/ Manager/ Owner.	21

In this table the similarity parameter of the category “U.S. Presidents” is 22 % only. It is because the category “U.S. Presidents” was recognized with a maximum of the similarity parameter 19.376 % in a case of Database E and model M72. Thus George W Bush looks similar to 41 U.S. Presidents with a relative probability $22/19.376= 1.135$ or 113.5 %. He also looks similar to other categories from this table, for instance, to the category “Taxi driver”. But in this case a relative probability is 74.7 % and 3 records only (“Taxi driver” is an unrepresentative category in this database). The similarity portrait could be used for a prediction of the social status of native. To increase a probability of this prediction several algorithms have been developed and verified [3, 7–8]. Test examples demonstrate the effectiveness of the system for the recognition of chats of respondents.

References

1. E.V. Lutsenko Conceptual principles of the system (emergent) information theory & its application for the cognitive modelling of the active objects (entities). 2002 IEEE International Conference on Artificial Intelligence System (ICAIS 2002). – Computer society, IEEE, Los Alamos, California, Washington – Brussels – Tokyo, p. 268–269. <http://csdl2.computer.org/comp/proceedings/icais/2002/1733/00/17330268.pdf>
2. Patent 2003610986, Russia, E.V. Lutsenko. Universal Cognitive Analytical System "AIDOS". Application № 2003610510, April 22, 2003.
3. 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.
4. Lois Rodden's AstroDatabank/www.astrodatabank.com
5. Richard Smoot. AstroDatabank, v. 4.00. Quick Start Guide.
6. E.V. Lutsenko, A.P. Trunev, V.N. Shashin. Typification and Identification of the Social Status of Respondents Based on the Astronomical Data at the Time of Birth. *Scientific Journal of the Kuban State Agricultural University*, No25 (1), 2007.
7. E.V. Lutsenko, A.P. Trunev. AST and Spectral Analysis of the Personal Information Using the Semantic Information Multi-Models. *Scientific Journal of the Kuban State Agricultural University*, No35 (1), 2008.
8. E.V. Lutsenko, A.P. Trunev. Increasing of the Personal Information Spectral Analysis Adequateness by AST Dividing on Typical and Untypical Parts. *Scientific Journal of the Kuban State Agricultural University*, No36 (2), 2008.