

УДК 57.087.1:004.931

ВЫДЕЛЕНИЕ ИНФОРМАТИВНЫХ ХАРАКТЕРИСТИК БИОЛОГИЧЕСКОГО ОБЪЕКТА ИЗ РЕГИСТРИРУЕМОГО СИГНАЛА

Артеменко М.В., Сотникова А.Д.

ФГБОУ ВО Юго-западный государственный университет, Курск, e-mail: artem1962@mail.ru

В работе рассматриваются перспективные методы формирования множества прямых и латентных информативных характеристик при анализе регистрируемых сигналов о состоянии биологического объекта в процессе мониторинга, представленных в виде временных трендов или таблиц «признак - значение». Предлагается применять: концепции системного анализа и самоорганизационного моделирования, анализ матрицы иерархий, синхронное математическое моделирование упреждающих и прогнозируемых значений сигнала от биологического объекта. В качестве способов формирования множества информативных характеристик биообъекта рассматривается: методика определения весовых коэффициентов информативных характеристик при вычислении латентных показателей, интегрально отражающих расстояния до центра определенного исследователем базового класса нормированных по дисперсии и основанная на вычислении критерия Стюдента; специфический способ формирования матрицы предпочтительности; применение нелинейной дискриминантной функции, идентифицируемой на основе полинома Колмогорова-Габор и метода группового учета аргументов (МГУА) и анализа рангов; критерий относительной погрешности отличий; критерий качества классификации в гипер-пространстве; применение опорных импульсов. Исследования в предлагаемых направлениях представляются перспективными для решения задач классификации или диагностики состояния анализируемого биообъекта.

Ключевые слова: биологический объект, информативные характеристики, матрица предпочтительности, самоорганизационное моделирование, анализ сигнала

ALLOCATION OF INFORMATIVE CHARACTERISTICS OF A BIOLOGICAL OBJECT FROM THE REGISTERED SIGNAL

Artemenko M. V., Sotnikov A. D.

Southwest state university, Kursk, email: artem1962@mail.ru

In work perspective methods of formation of a set of straight lines and latent informative characteristics are considered in the analysis of the registered signals of a condition of a biological object in the course of monitoring presented in the form of temporary trends or the tables "sign-value". It is offered to apply: concepts of the system analysis and self-organizational modeling, the analysis of a matrix of hierarchies, synchronous mathematical modeling of the anticipating and predicted values of a signal from a biological object. As ways of formation of a set of informative characteristics of a bioobject is considered: a technique of determination of weight coefficients of informative characteristics at calculation of the latent indicators which are integrally reflecting distances to the center of the basic class determined by the researcher rated on dispersion and Student's t-test based on calculation of criterion; specific way of formation of a matrix of preference; application of the nonlinear discriminant function identified on the basis of Kolmogorov-Gabor's polynom and the method of the group method of data handling (GMDH) and the analysis of ranks; criterion of a relative error of differences; criterion of quality of classification in hyper-space; application of basic impulses. Researches in the offered directions are represented perspective for the solution of problems of classification or diagnostics of a condition of the analyzed bioobject.

Keywords: biological object, informative characteristics, preference matrix, self-organizational modeling, analysis of a signal

Актуальной проблемой проектирования систем поддержки принятия классификационных решений соотнесения состояния биологического объекта к определенной страте (классу) является задача выделения маломощного сигнала на фоне большой зашумленности и характеристик с приемлемым уровнем адекватности значений диагностирующих состояние. Это касается и сложности выделения маломощных сигналов из шумов и формирования информативных множеств, достоверно характеризующих биообъект. С целью повышения эффективности диагностики заболеваний на ранних стадиях и улучшения качества медицинского обслуживания населения в целом.

Одним из наиболее распространенным подходом к построению классификацион-

ных систем является распознавание образов, начальным этапом которого является формирование множества информативных характеристик, характеризующих альтернативные классы. В дальнейшем, при построении систем дифференциальной диагностики (например, в медицине [18, 20]), что характерно для иерархических деревьев принятия решений, достаточно на каждом этапе бинарного дерева осуществить соотнесение объекта к одному из альтернативных классов w_0 и w_1 .

Для выбора множества характеристик, обладающих необходимой степенью классификационных возможностей, предлагается следующая методика. Пусть задано некоторое исходное множество характеристик $\{Pr\}_g$, где g – исходное количество, $g \neq 0$.

Требуется сформировать множество характеристик $\{Pr\}_g$, $1 \leq g \leq g$, обладающее наибольшими классификационными возможностями. Из разницы между множествами $\{Pr\}_g$ и $\{Pro\}_g$ возникает задача отбора характеристик, обладающих наибольшими классификационными возможностями, апробированными, статистически доказуемыми, формализованными методами.

Для этого необходимо иметь по каждому из альтернативных классов некоторые множества численных данных. Речь идет, в данном случае, о разведочном анализе и априорной принадлежности к нормальному закону распределения (в этом случае приемлемые результаты могут быть получены на выборках малого объема).

Наряду с регистрируемыми характеристиками (множеством $\{X\}$) для повышения качества классификации предлагается формировать дополнительное множество интегральных (множество $\{Y\}$), расчет элементов которого проводится по формуле (1).

$$Y_k = \frac{\sum_{i=1}^n (\alpha_{i,k} \cdot \frac{(x_i - \overline{Xw0}_i)^2}{\sigma w0_i})}{\sum_{i=1}^n \alpha_{i,k}} \quad (1),$$

Y_k - интегральная характеристика, k - индекс интегральной характеристики, $\overline{Xw0}_i, \sigma w0_i$ - соответственно, модальное значение (в случае нормального распределения совпадает с математическим ожиданием и средним значениями) и средне-квадратичное отклонение (СКО), полученное на обучающей выборке для некоторого класса «базовый», $\alpha_{i,k} \in [0,1]$ - весовые коэффициенты, определяющие информационный вклад характеристики x_i в формирование интегральной Y_k .

Весовые коэффициенты $\alpha_{i,k} \in [0,1]$ в указанной формуле предлагается определять следующими способами:

1. Путем назначения экспертами, исходя из личного опыта и анализа информационных источников.

2. Автоматически - на основе применения определенного статистически обоснованного математического аппарата.

3. Смешанным способом.

Во всех трех способах необходимо иметь единые шкалы измерений и ограничений. Принимаем ограничения:

$$\alpha_{i,k} \geq 0, \quad \sum_{i=1}^n \alpha_{i,k} = 1 \quad (2),$$

где n - количество частных показателей множества $\{X\}$, k - номер интегральной характеристики Y ($k \geq 0$).

Предлагается следующая методика определения рассматриваемых весовых коэффициентов $\alpha_{i,k}$ информативности дей-

ствительна только при наличии двух альтернативных классов):

1. Регистрируются для каждого альтернативного класса значения элементов множеств $\{X\}_n^{w_l} / l=0,1$ (где n - количество характеристик).

2. По значениям коэффициента Стьюдента на заданном уровне значимости (ошибки первого рода) осуществляется селекция с предположительно незначительными классификационными возможностями. В результате селекции для дальнейшего формирования информативного признакового пространства используется множество $\{X\}_{m_x}^{w_l}$, где m_x - количество характеристик, оставшихся после селекции.

3. По каждому показателю (признаку) $\{X\}_{m_x}^{w_l}$ вычисляются коэффициенты Стьюдента различий между двумя альтернативными классами - формируется множество $\{T\}_{m_x}$, элементами которого ($t_i / i=1, m_x$) являются значения коэффициентов Стьюдента.

4. Элементы полученного множества $\{T\}_{m_x}$ ранжируются (по убыванию или возрастанию). Строится диаграмма значений полученного ряда, по которой исследователь определяет необходимое количество ($m_y, m_y \leq m_x$) и «состав» интегральных характеристик путем выделения кластеров близких (по некоторой мере исследователя) значений ранжированных коэффициентов Стьюдента.

5. Каждому кластеру k ($k=1, \dots, m_y$) ставится в соответствие определенная интегральная характеристика Y_k и соответствующее подмножество $\{X\}_{m_x}^{w_l}$. Формула (2) модифицируется в формулу:

$$\alpha_{i,k} \geq 0, \quad \sum_{i=1}^{m_k} \alpha_{i,k} = 1 \quad (3).$$

6. Для каждой интегральной характеристики Y_k с учетом вычисленных в п.2 значений коэффициентов Стьюдента и выделенных в п.3 кластеров определяются значения соответствующих весовых коэффициентов в (3) по формуле:

$$\alpha_{i,k} = \frac{t_{i,k}}{\sum_{i=1}^{m_k} t_{i,k}} \quad (4).$$

Полученные величины весовых коэффициентов позволяют в дальнейшем формировать значения множества $\{Y\}$.

Таким образом, множество информативных характеристик $\{Pr\}_g$ для решения задачи обучения системы классификации (диагностики) формируется по формуле:

$$\{Pr\}_g = \{X\}_{m_x}^* + \{Y\}_{m_y}^* + \{X\}_{m_x}^* \quad (5),$$

где $\{X\}_{m_x}^*, \{Y\}_{m_y}^*, \{X\}_{m_x}^*$ - соответственно: множество отобранных характеристик по изложенной выше методике, множество

отобранных интегральных характеристик, множество отобранных экспертом характеристик, - элементами всех множеств являются «идентификаторы».

Для формирования информативного множества и принятия решения о включении в него признака так же рекомендуется использовать методологию принятия решений Саати Т.Л. [13]

Формируется матрица предпочтительности W , элементы которой для показателей i и j различаются по 9 степеням (признак i предпочтительнее j): $w_{ij}=1$ - равная степень предпочтительности, $w_{ij}=2$ - слабая $w_{ij}=3$ - средняя, $w_{ij}=4$ - предпочтение выше среднего, $w_{ij}=5$ - умеренно сильное, $w_{ij}=6$ - сильное, $w_{ij}=7$ - очень сильное (очевидное), $w_{ij}=8$ - очень, очень сильное (абсолютное), $w_{ij}=9$ - абсолютное.

Анализ матрицы позволяет после преобразования группировать признаки по кластерам предпочтительности с помощью предлагаемого П-преобразования. Строки I меняются местами со строками J до тех пор, пока сумма произведений значений элемента матрицы W^* на расстояние этого элемента до главной диагонали не станет минимальной (формула (6)). Таким образом, вокруг главной диагонали модифицированной матрицы предпочтений W^* сгруппируются элементы с максимальными значениями.

$$\left(\sum_{i=1}^N \sum_{j=1}^N w_{i,j}^* - |j-i| \right) \rightarrow \min, \quad (6)$$

$$w_{i,j}^* = \begin{cases} w_{i,j} & \text{для } i < j, \\ 10 - w_{i,j} & \text{для } i > j, \end{cases}$$

где N - количество анализируемых признаков до селекции.

Упорядочивание признаков по убыванию рангов информативности позволит определить степень предпочтения.

Целесообразно полагать, что для минимизации ошибки и принятия ложных решений из уже сформированного множества следует выделить ряд наиболее информативных характеристик, которые бы наиболее достоверно охарактеризовали исследуемый биологический объект.

Например, информативность отдельной характеристики из множества $\{X\}$ в работах [6, 7, 19] предлагается определять следующими методами [15].

Первый метод основан на нелинейной дискриминантной функции, определённой для классов (подразумевается бинарное иерархическое дерево решений). Он подразумевает:

- задание функции отклика для каждого класса w_0 и w_j ;

- структурно-параметрическую идентификацию полинома Габора, с помощью алгоритма МГУА [6, 11];

- вычисление мультипликанта для каждого аргумента;

- определение аддитивно-мультипликативного влияния показателя x_i на функцию отклика;

- введение «относительной погрешности отличий» (ОПО) (рекомендуется $0,01 \leq \varepsilon < 0,1$) и пересчёт значений величины мультипликативного влияния с учётом ОПО;

- упорядочивание x_i по мере убывания в диапазоне, заданным ОПО;

- формирование кортежей признаков для классов: $\{X^{w_0}\}$ и $\{X^{w_j}\}$;

- формирование множества рангов $R_n^{w_0}$ и $R_n^{w_j}$ по кортежам;

- формирование итогового множества информативных признаков;

- вычисление информативности признака $Inf(x_j)$ по формуле:

$$Inf(x_j) = \frac{\max(Rn_{x_j}^{w_0}, Rn_{x_j}^{w_j})}{\max_{j=1, N1}(\max(Rn_{x_j}^{w_0}, Rn_{x_j}^{w_j}))} \quad (7)$$

где $(Rn_{x_j}^{w_0}, Rn_{x_j}^{w_j})$ - значение ранга показателя x_j в w_0 и w_j , соответственно.

Во втором методе формирование множества информативных характеристик и вычисление значения $Inf(x_j)$, основывается на предварительной идентификации аппроксимирующего полинома Габора для каждой характеристики из начального множества $\{X\}$. В этом случае, процедура идентификации повторяется N раз для каждого класса w_0 и w_j , последовательно формируя множества $\{Z_j\} = \{X\} - x_j$ и отклики $Y(Z) = x_j$.

В результате формируются множества аппроксимантов для каждого класса. Аппроксиманты, значения коэффициента детерминации которых меньше определённого порогового значения, не учитываются. Минимальный объём множества аппроксимантов устанавливается исследователем - рекомендуется не менее 3.

Для каждого альтернативного класса формируются матрицы значений $(ApX)_{M0,N}^{w_0}$, по которым строятся векторы значений $(SapX)_{,N}^{w_0}$ и $(SapX)_{,N}^{w_j}$.

Кортежи показателей $\{X^{w_0}\}$ и $\{X^{w_j}\}$ формируются для каждого класса по мере убывания x_j по вектору значений с учётом ОПО.

Далее вычисление информативности осуществляется аналогично действиям, рассмотренным в способе 1.

В третьем методе переменные принимают значения «true» («1») либо «false» (0). С определённой точностью для возможности применения подходов, изложенных в способе 1 и 2, используют аналог полинома Габо-

ра $Y_{\sigma^*}(Z_{\sigma^*})$ для логических функций в виде формулы (8), на основании аналогов арифметических операций логическим функциям [10].

$$Y_{\sigma^*}(Z_{\sigma^*}) = 1 - \prod_{i=1}^N (1 - zb_i^*)^{p_k}, \quad p_k = \{0,1\}, \quad zb_i^* = \{0,1\}, \quad (8)$$

Далее расчеты аналогичны 1 и 2 методам.

В четвёртом методе осуществляется упорядочивание характеристик с последующим вычислением рангов, включением в информативный кортеж и вычислением информативности аналогично ранее рассмотренными способами на основании гиперобъемов H . При этом проводятся процедуры разведочной кластеризации с вычислением величины изменения качества кластеризации dH_{xy} при исключении из рассмотрения анализируемой характеристики [12].

dH_{xy} может принимать как положительные, так и отрицательные значения. В последнем случае после селекции наблюдается ухудшение качества классификации согласно общему гипер-объему пространства анализируемых показателей H .

Описанные методы применимы при априорно известном множестве характеристик. Однако на практике, например при исследовании биообъекта [2], часто необходимо определять информативность характеристики в текущий момент времени.

В случае, если биологический объект характеризуется некоторым сигналом небольшим объемом в работе [14] предлагается применить следующий подход.

В анализе сигнала используется фильтр, основанный на различии фаз измеряемого сигнала (регулярная функция) и шума (случайная функция, причём математическое ожидание шума принимается равным нулю).

Опорный импульс, применяемый в фильтре, имеет две прямоугольные ступени с различными амплитудами, общая длительность ступеней значительно меньше интервала наблюдения T_n .

При проведении измерений фрагмент искомого сигнала помещается на вторую ступень опорного импульса. В качестве измеряемой характеристики вычисляется фаза основной гармоники $e^{-jw\tau}$ при $w=1$. Для оценки величины сигнала используется разность фаз опорного импульса с сигналом и без него.

При значениях D и M ($M=T_H/T_{HP}$ а $D=U_f/U_j$), близких к единице, увеличивается чувствительность фазы к изменению напряжения (амплитуды ступени).

При этом зависимость фазы от D и M имеет вид [1, 15]:

$$\phi(w) = \arctg \left[\frac{1-D}{1+D} \operatorname{tg} \left(\frac{\pi}{2M} \right) \right] \quad (9)$$

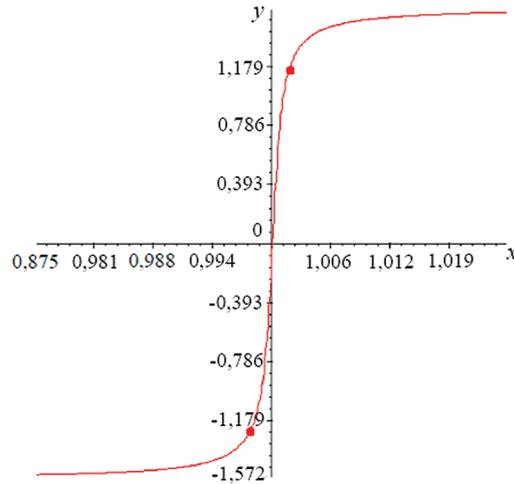


Рис. 1. Зависимость значения фазы основной гармоники опорного сигнала от отношения амплитуд первой и второй ступеней

Для повышения информативности результатов следует обеспечивать попадание фазы опорного импульса и фазы импульса с изменённым фрагментом сигнала на квазилинейный участок графика зависимости от параметра $\phi(D)$ (точки на кривой рис. 1). Для этого первую ступень опорного импульса следует принять равной 1, затем вычислить значение D (начальная фазы основной гармоники ($w=1$) попадёт на нижний конец квазилинейного участка кривой $\phi(D)$).

$$\phi'(D)_{\max} = \frac{-\sin \alpha}{\cos 2\alpha} \quad (10)$$

После формирования данным образом опорного импульса вычисляется среднее значение измеряемого фрагмента сигнала, которое может быть оценено через разность фаз основной гармоники опорного импульса с сигналом и без него [1].

Затем этот фрагмент умножается на коэффициент:

$$\alpha = \frac{2(1-D)}{\bar{x}}, \quad (11)$$

где \bar{x} - среднее значение фрагмента «сигнал + шум».

Восстановление сложных сигналов проводят путём оценки средних значений отдельных их фрагментов. Для этого исходный сигнал формируется в виде импульса с колоколовидной огибающей, к нему добавляется шум в виде последовательности случайных нормально распределённых чисел. Зашумлённый сигнал разбивается на фрагменты, для каждого такого фрагмента описанным выше способом оценивается среднее значение сигнала [15].

В результате из некогерентного шума формируется информативный сигнал, по которому можно вести дальнейший анализ [9].

Для вычисления значения информативных параметров снятого сигнала [5,17], предлагается использовать функциональные параметры, характеризующие поведение системы, способной принимать решение в текущий момент времени в соответствии с учетом прошлого опыта и прогнозирования будущего [3, 4].

Целесообразно полагать, что биологическая система должна принимать решения в момент времени t , при регистрации значения информативного сигнала $y(t)$. Так как автономная система управления организмом систематически учитывает лучший результат «прошлого» опыта и автоматически прогнозирует развитие ситуации в будущем, то сигнал $y(t)$ может быть представлен в виде [8]:

$$\tilde{y}(t) = F(\phi_{t-1}(t-1), \phi_{t-2}(t-2), \dots, \phi_{t-T_1}(t-T_1), \phi_{t+1}(t+1), \phi_{t+2}(t+2), \dots, \phi_{t+T_2}(t+T_2)), \quad (12)$$

где: $\phi_{t\pm i}(t\pm i)$ – частный аргумент-функционал, отражающий состояние анализируемой системы в моменты времени $(t\pm i)$; T_1 и T_2 – максимальное время упреждения и экстраполяции (прогноза), соответственно; T_1 и T_2 формирующий индекс частных аргумент-функционалов, $F()$ интегральный «опытно-прогностический функционал» – FPF(t).

В первом приближении частные спектральные аргумент-функционалы представляются линейными преобразованиями:

$$\phi_{t\pm i}(t\pm i) = a_{t\pm i} + b_{t\pm i} \cdot y(t\pm i). \quad (13)$$

Множество значений информативных хронометрических параметров в момент времени $t+I$, формируется «Модулем «фантазий», который содержит правила идентификации $\{Y\}_{t+1}^f$ по $\{Y\}_t$ и рассогласования между прогнозируемым состоянием S_{t+1} и реально регистрируемым $\{S\}_{t+1}^r$. Обсуждение заканчивается, когда указанное рассогласование достигнет приемлемого порога величины [8] (задается в общем случае нижней и верхними границами).

Выводы

Рассмотренные подходы позволяют формализовать выделение множеств информативных прямых и латентных характеристик биологического объекта путём анализа сигналов различных уровней мощности на основе применения как упреждающих, так и прогнозируемых значений сигнала. Анализ осуществляется с помощью применения самоорганизационной концепции моделирования, дифференцированного анализа и метода анализа иерархий. Исследования в этом направлении представляются новыми и пер-

спективными для решения задачи диагностики состояния анализируемого биообъекта.

Список литературы

1. Авдеева Д. К., Вылегжанин О. Н., Рыбалка С. А. Метод выделения полезного сигнала при помехе // Успехи современного естествознания. – 2010. - №1. – С. 132-135.
2. Артёмченко Н.М. Распознавание состояний легких человека по издаваемому ими акустическому шуму // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2015. № 2 (15). стр. 94-98.
3. Артеменко М.В., Бабков А.С. Классификация методов прогнозирования поведения систем // Современные проблемы науки и образования. – 2013. – № 6.; URL: <http://www.science-education.ru/ru/article/view?id=11527> (дата обращения: 10.12.2016)
4. Артеменко М.В., Добровольский И.И., Мишустин В.Н. Информационно-аналитическая поддержка автоматизированной классификации на основе прямых и обратных решающих правил на примере прогноза тромбозов. // Современные наукоемкие технологии. – 2015. №12 (часть 2). – стр.199-205.
5. Артеменко М.В., Калугина Н.М. Диагностический анализ состояния биообъекта по хронометрическим параметрам регистрируемых сигналов. //Международный научно-исследовательский журнал. 2016. № 4-2 (46). стр. 30-35.
6. Артеменко М. В., Калугина Н. М., Шуткин А. Н. Формирование множества информативных показателей на основании аппроксимирующего полинома Колмогорова-Габор и максимального градиента функциональных различий // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение, 2016, №1, стр. 116-223
7. Артеменко М.В., Подвальный Е.С., Старцев Е.А. Метод комплексной оценки и выборка состава информативных признаков в задачах оценки состояния биотехнических систем //Биомедицинская радиоэлектроника, 2016, №9, стр.38-44
8. Калугина Н. М. Система диагностики состояния организма по спектральным характеристикам биомедицинского сигнала. //Лучшая статья 2016: сборник статей III Международного научно-практического конкурса / Под общ. ред. Г. Ю. Гуляева – Пенза: МЦНС «Наука и Просвещение». – 2016. – 200 с.
9. Кобринский В.А Телемедицина в системе практического здравоохранения. -М.: Директ-Медиа, 2016. – 238 с.
10. Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопрос статистической устойчивости решений Новосибирск: Изд-во Ин-та математики, 1999. 212 с.
11. Орлов А.А. Принципы построения архитектуры программной платформы для реализации алгоритмов метода группового учета аргументов / А.А. Орлов // Управляющие системы и машины.-2013. - №2. -стр.65-71
12. Превентивная медицина: Опыт работы информационного полипараметрического комплекса / Под ред. Н.В. Дмитриевой. –М.: Книжный дом «ЛИБРОКОМ», 2010. – 248 с.
13. Саати Томас Л. Принятие решений при зависимостях и обратных связях: аналитические сети. Пер. с англ./ Науч. ред. А.В. Андрейчиков, О.Н. Андрейчикова. Изд. 4-е. – М.: ЛЕНАНД, 2015. – 360 с.
14. Сотникова А. А., Писарев М. В., Калугина Н.М. Выделение информативных признаков //Медико-экологические информационные технологии – 2016: сборник научных статей по материалам XIX Международной научно-технической конференции / редкол.: Н. А. Корневский [и др.]; Юго-Зап. гос. ун-т. – Курск, 2016. – 308 с.
15. Способ измерения сигналов произвольной формы в присутствии случайных шумов/Д.К. Авдеева. Пат. 2133474 РФ, МКИ 19/02. Заявл. 30.10.97; Оpubл. 20.07.99, Бюл. №20.
16. Справочник по типовым программам моделирования / А.Г. Ивахненко, Ю.В. Коппа, В.С. Степашко и др.; под ред. Ивахненко А.Г. – К.: Техника,1980.- 184 с.
17. Цифровая обработка сигналов и изображений в радиофизических приложениях. / Под ред. В. Ф. Кравченко, - М.:ФИЗМАТЛИТ, 2007. -544 с.
18. Р. Хэгглин. Дифференциальная диагностика внутренних болезней –М.: Триада – X, 2015, -800 с.
19. Artemenko M.V., Kalugina N.M., Dobrovolsky I.I. The formation of a set of informative features based on the functional relationships between the data structure field observations // European Journal of Natural History. 2016. № 6. С. 43-48.
20. Differential Diagnosis: Methodological Instructions for Students // URL: <http://zavantag.com/docs/index-16795051.html?page=13> (дата вхождения – 20.12.2016)