

УДК 004.8
06.02.00 Ветеринария и Зоотехния

АГЛОМЕРАТИВНАЯ КОГНИТИВНАЯ КЛАСТЕРИЗАЦИЯ НОЗОЛОГИЧЕСКИХ ОБРАЗОВ В ВЕТЕРИНАРИИ

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Scopus Author ID: 57191193316
РИНЦ SPIN-код: 9523-7101
prof.lutsenko@gmail.com <http://ic.kubagro.ru>
*Кубанский государственный аграрный университет
имени И.Т.Трубилина*

В статье на небольшом численном примере рассматривается сходство и различие нозологических образов в ветеринарии с применением нового метода агломеративной кластеризации, реализованного в автоматизированном системно-когнитивном анализе (АСК-анализ). Этот метод получил название: «Агломеративная когнитивная кластеризация». Этот метод отличается от известных традиционных тем, что: а) в нем параметры обобщенного образа кластера вычисляются не как средние от исходных объектов (классов) или их центр тяжести, а определяются с помощью той же самой базовой когнитивной операции АСК-анализа, которая применяется и для формирования обобщенных образов классов на основе примеров объектов и которая действительно корректно обеспечивает обобщение; б) в качестве критерия сходства используется не евклидово расстояние или его варианты, а интегральный критерий неметрической природы: «суммарное количество информации», применение которого теоретически корректно и дает хорошие результаты в неортономированных пространствах, которые обычно и встречаются на практике; в) кластерный анализ проводится не на основе исходных переменных, матриц частот или матрицы сходства (различий), зависящих от единиц измерения по осям, а в когнитивном пространстве, в котором по всем осям (описательным шкалам) используется одна единица измерения: количество информации, и поэтому результаты кластеризации не зависят от исходных единиц измерения признаков объектов. Все это позволяет получить результаты кластеризации, понятные специалистам и поддающиеся содержательной интерпретации, хорошо согласующиеся с оценками экспертов, их опытом и интуитивными ожиданиями, что часто представляет собой проблему для классических методов кластеризации

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС», КОГНИТИВНОЕ ПРОСТРАНСТВО, АГЛОМЕРАТИВНАЯ КЛАСТЕРИЗАЦИЯ
Doi: 10.21515/1990-4665-138-033

UDC 004.8
Veterinary and Zootechnics

AGGLOMERATIVE COGNITIVE CLUSTERING OF NOSOLOGICAL IMAGES IN VETERINARY MEDICINE

Lutsenko Evgeny Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
Scopus Author ID: 57191193316
RSCI SPIN-code: 9523-7101
prof.lutsenko@gmail.com <http://ic.kubagro.ru>
Kuban State Agrarian University, Krasnodar, Russia

The article deals with the similarity and difference of nosological images in veterinary medicine using a new method of agglomerative clustering implemented in Automated system-cognitive analysis (ASC-analysis) on a small numerical example. This method is called Agglomerative cognitive clustering. This method differs from the known traditional facts: a) parameters of a generalized image of the cluster are computed not as averages from the original objects (classes) or their center of gravity, and are defined using the same underlying cognitive operations of ASC-analysis, which is used for the formation of generalized images of the classes on the basis of examples of objects and which is really correct and provides a synthesis; b) as a criterion of similarity we do not use Euclidean distance or its variants, and the integral criterion of non-metric nature: "the total amount of information", the use of which is theoretically correct and gives good results in non-orthonormal spaces, which are usually found in practice; c) cluster analysis is not based on the original variables, matrices of frequency or a matrix of similarities (differences) dependent on the measurement units of the axes, and in the cognitive space in which all the axes (descriptive scales) use the same unit of measurement: the quantity of information, and therefore, the clustering results do not depend on the original units of measurement features. All this makes it possible to obtain clustering results that are understandable to specialists and can be interpreted in a meaningful way that is in line with experts' assessments, their experience and intuitive expectations, which is often a problem for classical clustering methods

Keywords: AUTOMATIVE SYSTEM-COGNITIVE ANALYSIS, INTELLECTUAL SYSTEM "EIDOS", COGNITIVE SPACE, AGGLOMERATIVE CLUSTERIZATION

В данной статье процесс диагностики и лечения животных рассматривается с позиций теории управления с применением искусственного интеллекта.

Состав классической системы управления с обратной связью приведен на рисунке 1:

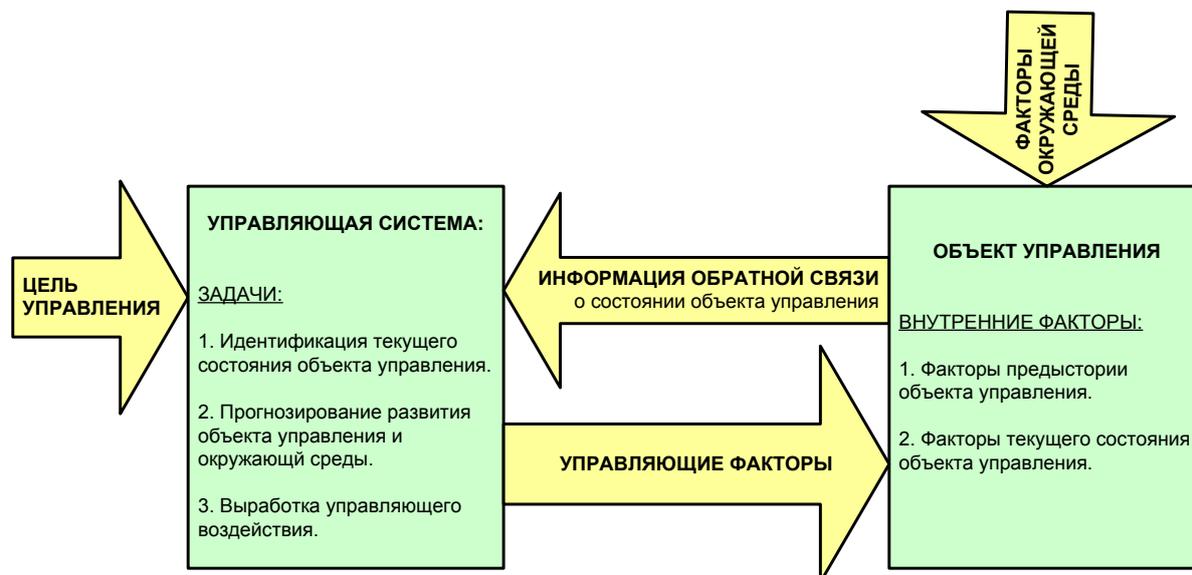


Рисунок 1. Цикл управления в замкнутой системе управления

Конкретизируем стандартные универсальные термины теории управления для ветеринарии:

- объектом управления является животное;
- управляющая система – это ветеринар;
- информация обратной связи о состоянии объекта управления – это клиническая картина (непосредственно наблюдаемая симптоматика) и, возможно, данные лабораторных исследований;
- управляющие факторы – это лечебные воздействия со стороны ветеринара на больное животное;
- целью управления является перевести животное из нежелательного состояния болезни в здоровое состояние.

Для принятия решения об управляющих воздействиях, т.е. для разработки *плана лечения*, необходимо предварительно поставить больному

животному *нозологический диагноз* (от греч. *posos* — болезнь и *logos* — слово, учение; букв.— учение о болезни), т.е. выработать заключение о сущности заболевания и состоянии животного, выраженное в нозологических терминах (установить болезненную форму или нозологическую единицу).

Задача нозологической диагностики может решаться в частности на основе знания симптоматики и по сути представляет собой задачу классификации объекта по его признакам (идентификация, распознавание). Решение этой задачи в ветеринарии (как и задач: поддержки управляющих решений и исследования моделируемой предметной области путем исследования ее модели) подробнейшим образом рассмотрено в работе [1]. Для решения этих задач применен автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий Универсальная автоматизированная система «Эйдос» [2, 3]¹.

В данной работе рассматривается и решается **проблема**, возникающая при нозологической диагностике (классификации), состоящая в том, что различные нозологические единицы в ряде случаев имеют сходную симптоматику и являются сложно различимыми из-за чего возникает опасность постановки неверного диагноза и, соответственно, неправильного лечения².

Предлагается идея решения проблемы состоящая в том, что ветеринар должен иметь информацию о степени сходства-различия нозологических образов по симптоматике должен учитывать ее при интерпретации выходных форм по результатам диагностики. Если говорить более конкретно, то ветеринар понимая *этиологию* различных заболеваний может видеть глубокую внутреннюю взаимосвязь между ними и иногда может

¹ http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf

² Известно, что лечение не от той болезни иногда может быть привести к более плачевным результатам, чем отсутствие лечения. Это может быть связано как с весьма нежелательными побочными эффектами применяемых методов, так и в очень большой степени просто с потерей ценнейшего времени.

увидеть причину симптоматически более ярко выраженного заболевания в другом заболевании, симптоматика которого выражена менее четко.

Отметим, что первоначально под нозологией понимали раздел патологии, включающий общее учение о болезни (общая нозология), а также изучающий этиологию, патогенез клинические и патологоанатомические проявления отдельных болезней (частная нозология), а также *классификацию* и номенклатуру болезней³.

В качестве примера рассмотрим результат нозологической диагностики, приведенный на рисунке 21 из работы [1], отображающий сходство клинической картины конкретного пациента («Мурка») с обобщенными нозологическими образами (рисунок 2):

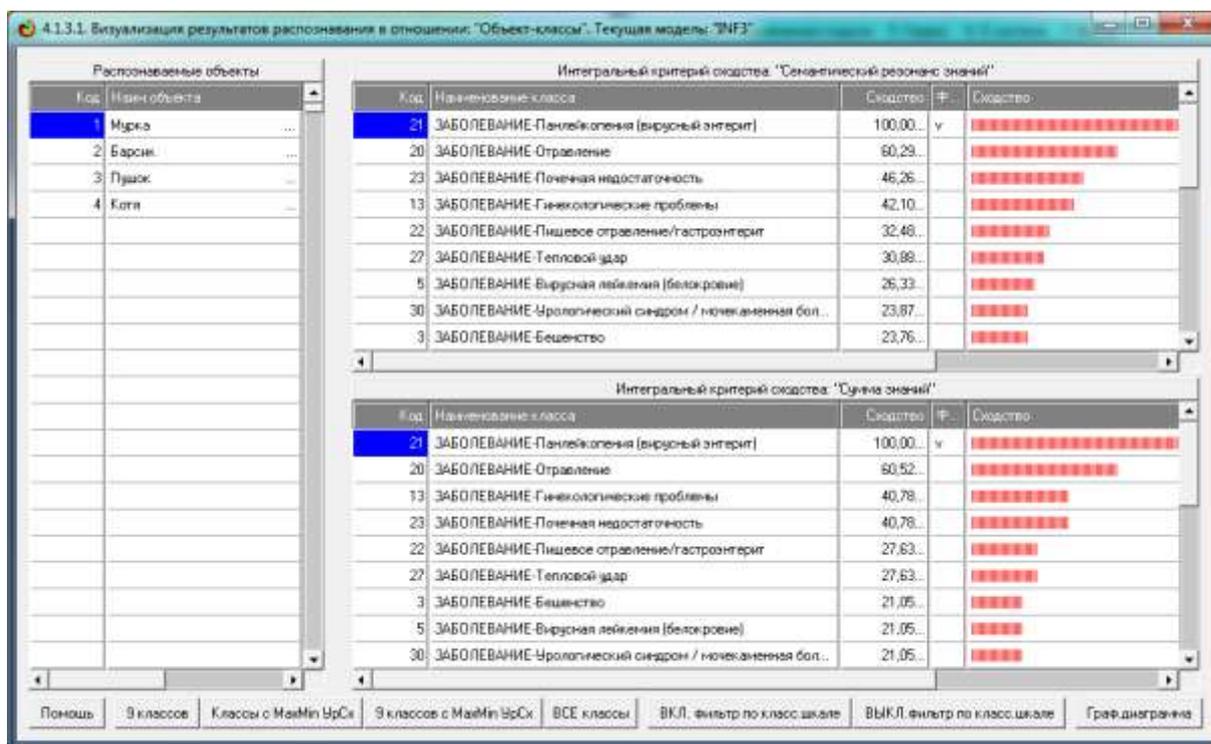


Рисунок 2. Экранная форма с результатами нозологической диагностики пациента «Мурка», (симптоматика из обучающей выборки)

Возникает вопрос, от какого заболевания лечить Мурку, от 1-го в этом списке, или 2 или может быть даже 3-го.

³ В этом понимании нозология не имеет чётких границ с патологией.

Для ответа на этот вопрос применим кластерный анализ нозологических образов.

Кластерный анализ (англ. *Data clustering*) – это задача разбиения заданной выборки *объектов* (ситуаций) на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

Однако мы не будем применять классические методы кластерного анализа и реализующий их программный инструментарий. Причиной этого является то, что эти методы имеют ряд довольно существенных на взгляд автора недостатков. Эти недостатки подробно проанализированы в работе [4], в ней же предложено и их решение [4, 5]. Поэтому в данной статье мы не будем их подробно рассматривать. Отметим лишь, что многие из этих недостатков, являющихся проблемами классических методов кластерного анализа, преодолены в АСК-анализе и интеллектуальной системе «Эйдос» (таблица 1).

Таблица 1 – ПРОБЛЕМЫ КЛАССИЧЕСКИХ МЕТОДОВ КЛАСТЕРИЗАЦИИ И ИХ РЕШЕНИЕ ПРЕДЛАГАЕМЫЕ В АСК-АНАЛИЗЕ И СИСТЕМЕ «ЭЙДОС»

№	Формулировка проблемы кластерного анализа	Предлагаемое в АСК-анализе и реализованное в системе «Эйдос» решение
1.	<i>Проблема 1.1</i> выбора метрики, корректной для неортонормированных пространств.	Предлагается применить неметрический интегральный критерий, представляющий собой суммарное количество информации в системе признаков о принадлежности объекта к классу («информационное расстояние»), никак не основанный на предположении об ортонормированности пространства и корректно работающий в неортонормированных пространствах
2.	<i>Проблема 1.2</i> ортонормирования пространства.	Поддерживается
3.	<i>Проблема 2.1</i> сопоставимой обработки описаний объектов, описанных признаками различной	Объекты формально описываются в виде векторов, затем рассчитывается матрица абсолютных частот и на ее ос-

	<i>природы, измеряемыми в различных единицах измерения (проблема размерностей).</i>	нове – матрица знаний, с использованием которой все признаки измеряются в одних единицах измерения: единицах измерения количества данных, информации и знаний – битах, байтах, и т.д.
	<i>Проблема 2.2 формализации описаний объектов, имеющих как количественные, так и качественные признаки.</i>	Числовые шкалы преобразуются в интервальные значения, после чего градации всех типов шкал обрабатываются единообразно (см.п.3)
4.	<i>Проблема 3.1 доказательства гипотезы о нормальности исходных данных.</i>	Нет необходимости, т.к. предлагаемые частные и интегральные критерии не предполагают нормальности исходных данных
5.	<i>Проблема 3.2 нормализации исходных данных.</i>	Реализованы режимы ремонта или взвешивания исходных данных.
6.	<i>Проблема 3.3 применения непараметрических методов кластеризации, корректно работающих с ненормализованными данными.</i>	Предлагаемые методы являются непараметрическими и корректно работают с ненормализованными данными
7.	<i>Проблема 4 разработки такого метода кластерного анализа, математическая модель и алгоритм и которого органично включали бы фильтр, подавляющий шум в исходных данных, в результате чего данный метод кластеризации корректно работал бы при наличии шума в исходных данных.</i>	Предлагаемый метод включает фильтр подавления шума на уровне формирования матрицы абсолютных частот и самой математической форме интегрального критерия. Кроме того, реализованы режимы удаления или корректной обработки артефактов, выбросов (нетипичных объектов) и малопредставленных данных, по которым нет достаточной статистики в исходных данных
8.	<i>Проблема 5 разработки метода кластерного анализа, математическая модель и алгоритм и которого обеспечивали бы <u>выявление</u> «выбросов» (артефактов) в исходных данных и позволяли либо вообще не показывать их в дендрограммах, либо показывать, но так, чтобы было наглядно видно, что это артефакты.</i>	Поддерживается исключение выбросов и артефактов из дендрограмм, либо их отображение специальным для них образом.

В статье [4] были отмечены и недостатки самого предложенного в ней решения, а также указаны перспективы его развития. Эти недостатки связаны прежде всего с программной реализацией DOS-версии самой сис-

темы «Эйдос», которая была текущей на момент написания работы [4]. В текущей версии системы «Эйдос» от 21.04.2018 все эти ограничения DOS-версии преодолены и реализованы многие из указанных в [4] перспектив. В частности реализован режим агломеративной когнитивной кластеризации (режим 4.2.2.3). Вместо описания данного режима приведем его Help (рисунок 3):

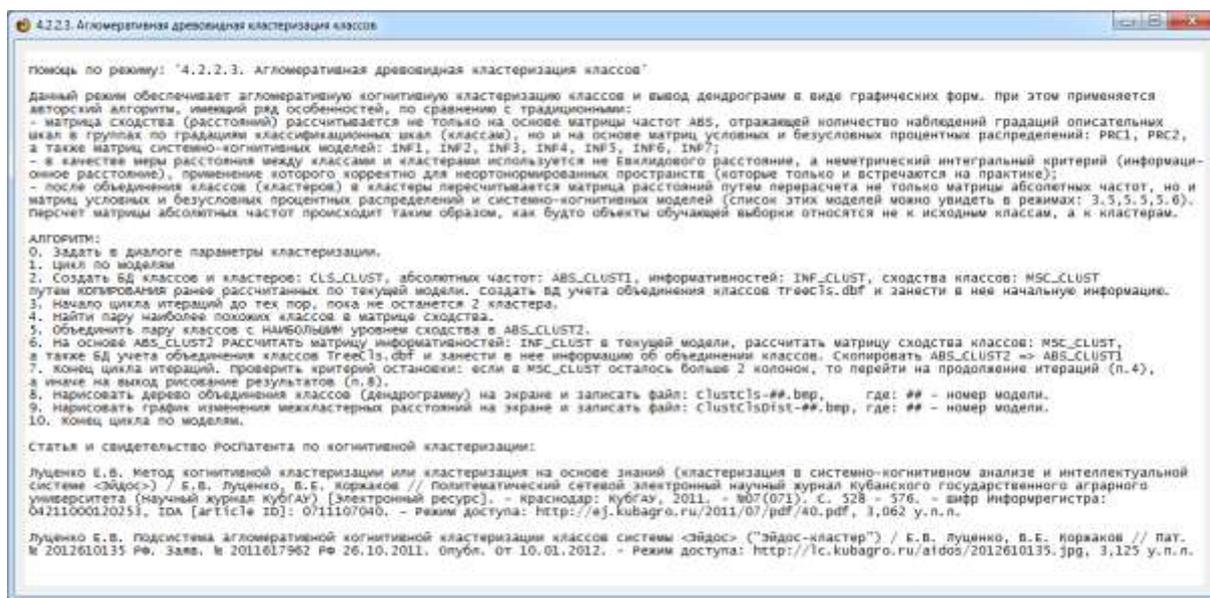
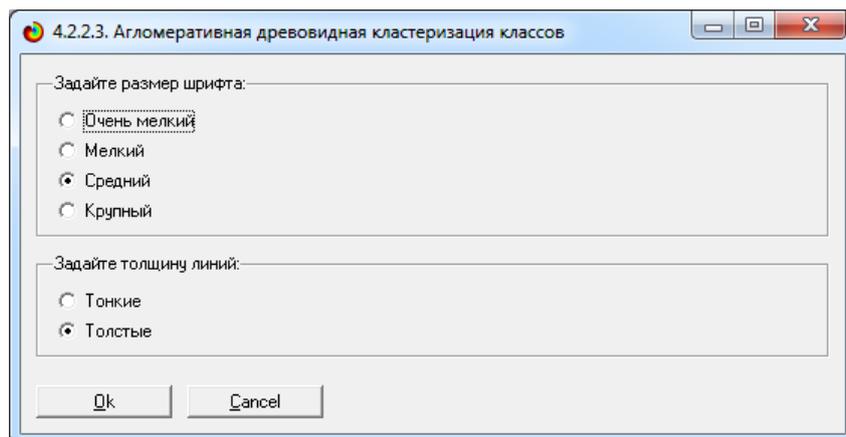
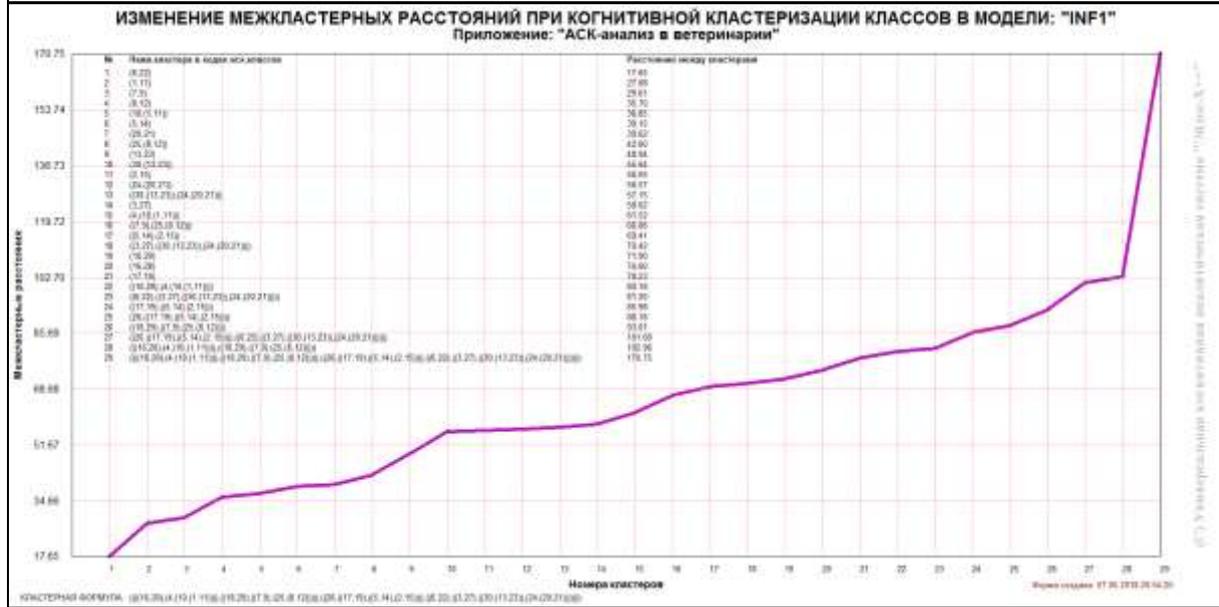
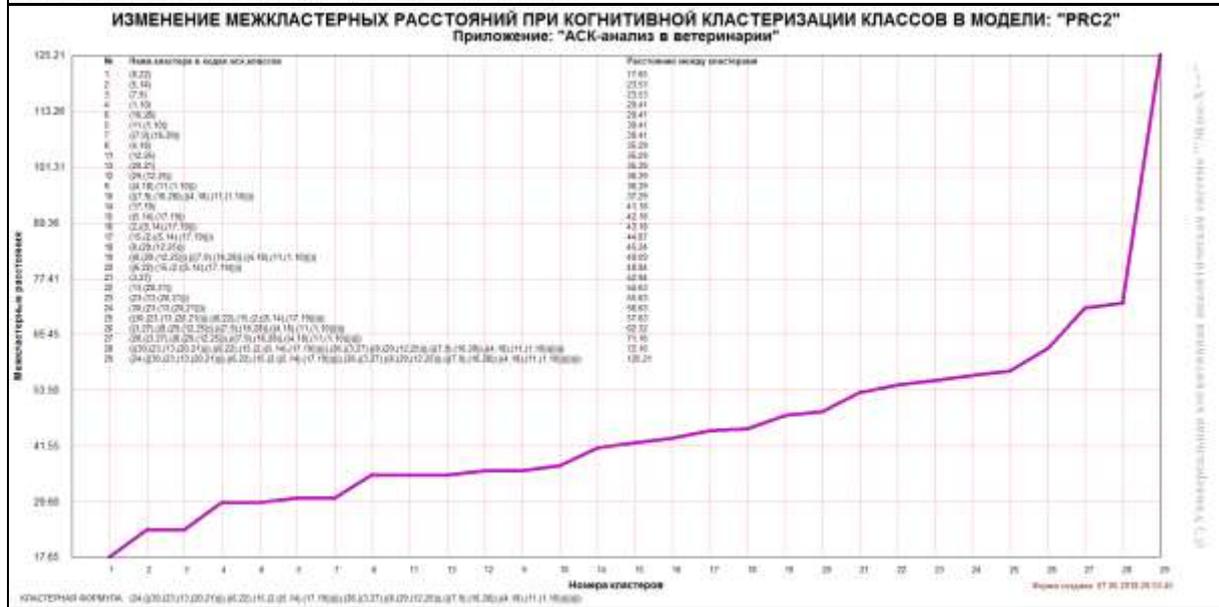
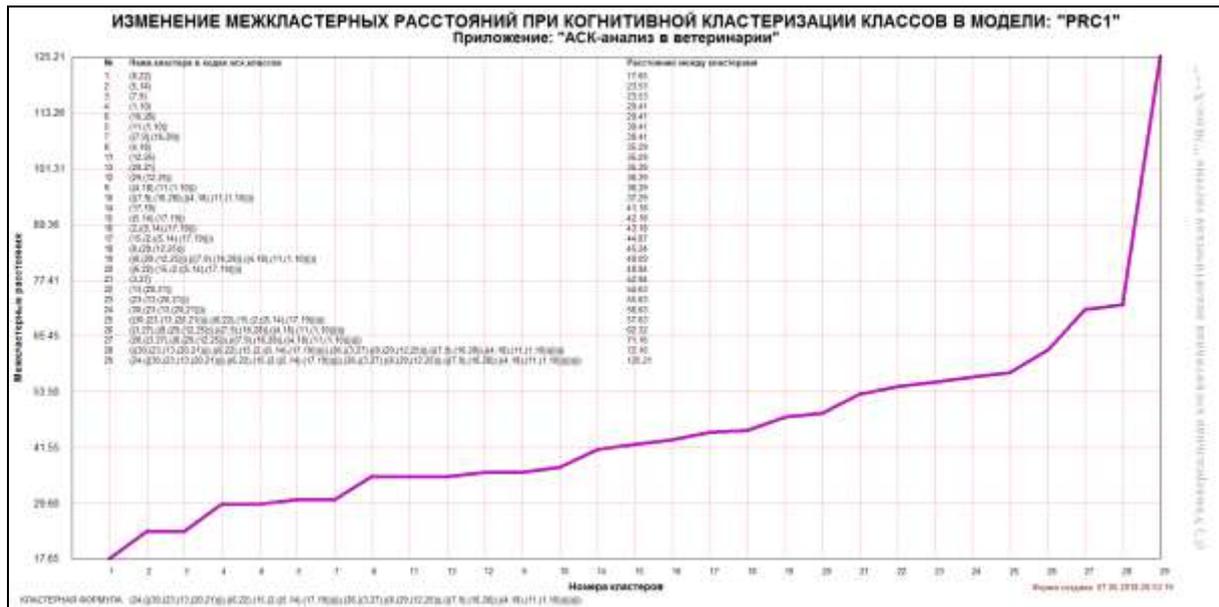


Рисунок 3. окно настроек визуализации дендрограмм и Help режима агломеративной когнитивной кластеризации

На рисунках 4 приведены агломеративные дендрограммы нозологических образов, построенные на основе моделей, созданных и исследованных в работе [1].



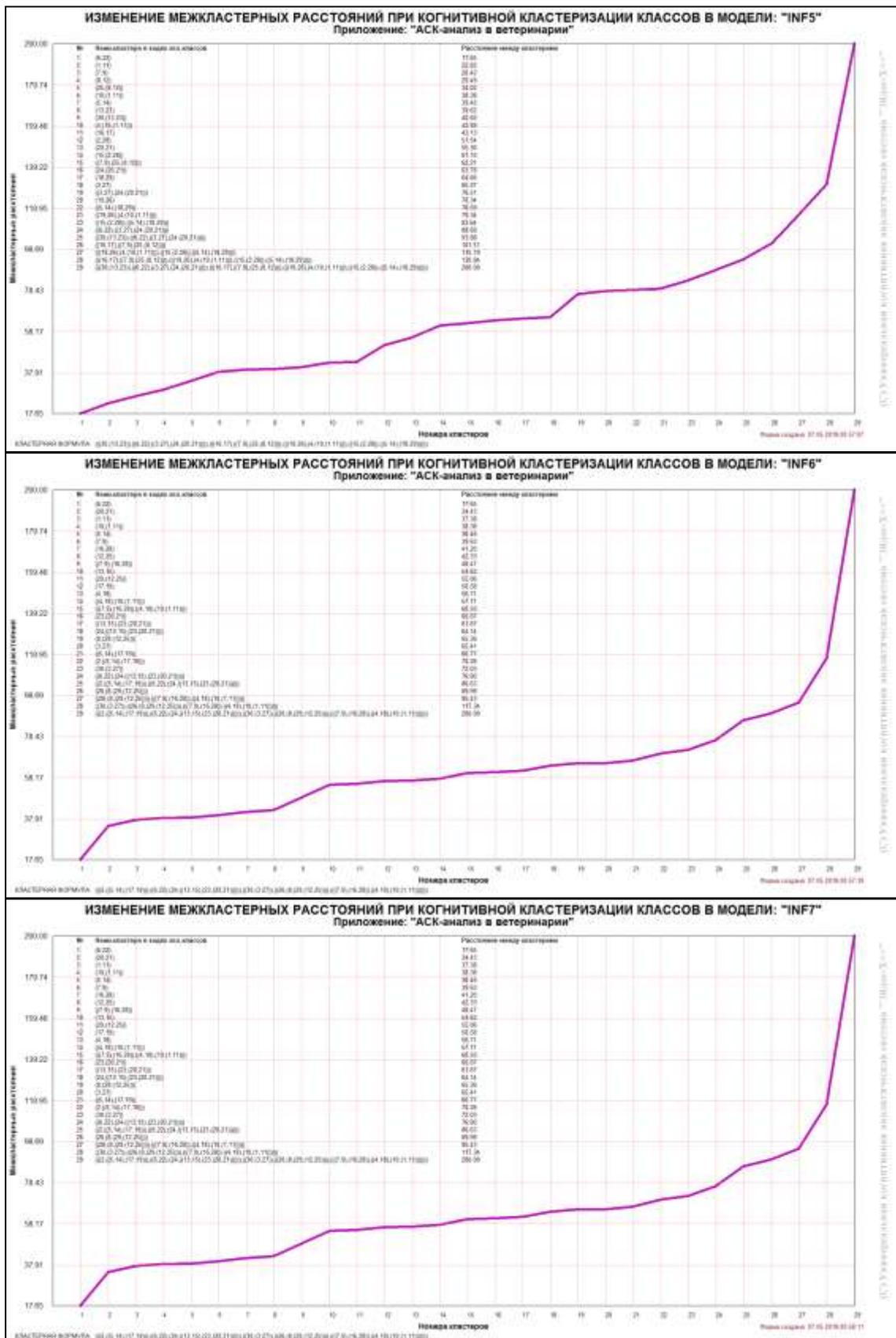


Рисунок 5. Межкластерные расстояния нозологических образов, построенные на основе моделей, созданных и исследованных в работе [1]

ватным средством решения поставленной проблемы, т.к. позволяет достоверно оценить сходство-различие нозологических образов в ветеринарии по их симптоматике. Это позволяет ветеринару более обоснованно применять решение о назначении лечения с учетом не только результатов диагностики, но и классификации самих нозологических образов и их этиологии.

В качестве **недостатка** предлагаемого метода можно указать на более значительные затраты времени и вычислительных ресурсов по сравнению с классическими традиционными методами кластеризации.

Как **перспективу** развития метода отметим: повышение размерности дендрограмм и разработку режима агломеративной когнитивной кластеризации признаков.

Материалы данной статьи могут быть использованы в учебном процессе при преподавании дисциплин: «Интеллектуальные информационные системы», «представление знаний в информационных системах», «Современные технологии в образовании (магистратура)», «Управление знаниями (магистратура)», «Введение в искусственный интеллект», «Интеллектуальные и нейросетевые технологии в образовании (магистратура)», «Основы искусственного интеллекта» и других.

Планируется продолжить исследования и разработки, представленные в данной статье, по следующим направлениям:

1. Агломеративная когнитивная кластеризация симптомов и синдромов в ветеринарии.

2. Разработка в среде интеллектуальной системы Эйдос-Х++ ветеринарных диагностических тестов и проведение научных исследований на основе эмпирических данных (данных наблюдений и экспериментов).

3. Интеграция в среде интеллектуальной системы Эйдос-Х++ без программирования нескольких диагностических тестов в батарею тестов или в супертест с одной общей моделью.

Этим и другим применениям способствует и то, что система «Эйдос» является мультязычной интеллектуальной on-line средой для обучения и научных исследований [3, 4]⁴ и находится в полном открытом бесплатном доступе (причем с подробно комментированными открытыми исходными текстами: <http://lc.kubagro.ru/AIDOS-X.txt>) на сайте автора по адресу: <http://lc.kubagro.ru/aidos/Aidos-X.htm>.

В заключение автор выражает благодарность проректору по научной работе Кубанского ГАУ им. И.Т. Трубилина доктору биологических наук профессору Андрею Георгиевичу Коцаеву за помощь в публикации статьи.

Список литературы

1. Луценко Е.В. Автоматизированный системно-когнитивный анализ в ветеринарии (на примере разработки диагностических тестов) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №03(137). С. 143 – 196. – IDA [article ID]: 1371803031. – Режим доступа: <http://ej.kubagro.ru/2018/03/pdf/31.pdf>, 3,375 у.п.л.
2. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная on-line среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 у.п.л.
3. Луценко Е.В., Открытая масштабируемая интерактивная интеллектуальная on-line среда «Эйдос» («Эйдос-online»). Свид. РосПатента РФ на программу для ЭВМ, Заявка № 2017618053 от 07.08.2017, Гос.рег.№ 2017661153, зарегистр. 04.10.2017. – Режим доступа: <http://lc.kubagro.ru/aidos/2017661153.jpg>, 2 у.п.л.
4. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.
5. Луценко Е.В., Подсистема агломеративной когнитивной кластеризации классов системы «Эйдос» ("Эйдос-кластер"). Пат. № 2012610135 РФ. Заяв. № 2011617962 РФ 26.10.2011. Оpubл. От 10.01.2012. – Режим доступа: <http://lc.kubagro.ru/aidos/2012610135.jpg>, 3,125 у.п.л.

⁴ http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf

Spisok literatury

1. Lutsenko E.V. Avtomatizirovannyi sistemno-kognitivnyy analiz v veterina-rii (na primere razrabotki diagnosticheskikh testov) / E.V. Lutsenko // Politematicheskiy setevoy elektronnyy nauchnyy zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyy zhurnal KubGAU) [Elektronnyy resurs]. – Krasnodar: Kub-GAU. 2018. – №03(137). S. 143 – 196. – IDA [article ID]: 1371803031. – Rezhim dostupa: <http://ej.kubagro.ru/2018/03/pdf/31.pdf>. 3.375 u.p.l.
2. Lutsenko E.V. Otkrytaya masshtabiruyemaya interaktivnaya intellektualnaya on-line sreda dlya obucheniya i nauchnykh issledovaniy na baze ASK-analiza i sistemy «Ey-dos» / E.V. Lutsenko // Politematicheskiy setevoy elektronnyy nauchnyy zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyy zhurnal KubGAU) [Elektronnyy resurs]. – Krasnodar: KubGAU. 2017. – №06(130). S. 1 – 55. – IDA [article ID]: 1301706001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>. 3.438 u.p.l.
3. Lutsenko E.V.. Otkrytaya masshtabiruyemaya interaktivnaya intellektualnaya on-line sreda «Eydos» («Eydos-online»). Svid. RosPatenta RF na programmu dlya EVM. Za-yavka № 2017618053 ot 07.08.2017. Gos.reg.№ 2017661153. zaregistr. 04.10.2017. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2017661153.jpg>. 2 u.p.l.
4. Lutsenko E.V. Metod kognitivnoy klasterizatsii ili klasterizatsiya na osnove znaniy (klasterizatsiya v sistemno-kognitivnom analize i intellektualnoy sisteme «Eydos») / E.V. Lutsenko. V.E. Korzhakov // Politematicheskiy setevoy elektronnyy nauchnyy zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyy zhurnal KubGAU) [Elektronnyy resurs]. – Krasnodar: KubGAU. 2011. – №07(071). S. 528 – 576. – Shifr Informregistra: 0421100012\0253. IDA [article ID]: 0711107040. – Rezhim dostupa: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>. 3.062 u.p.l.
5. Lutsenko E.V.. Podсистема aglomerativnoy kognitivnoy klasterizatsii klas-sov sistemy «Eydos» ("Eydos-klaster"). Pat. № 2012610135 RF. Zayav. № 2011617962 RF 26.10.2011. Opubl. Ot 10.01.2012. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2012610135.jpg>. 3.125 u.p.l.