12. Usenova, A.K.Osobennosti vodnyh problem v Central'noj Azii [Features of water problems in Central Asia] // Social'no-ekonomicheskie nauki i gumanitarnye issledovaniya[Socio-economic sciences and humanities research].- 2014. №3(22).– PP.174-179.
13. Ajmen, A.T. Gosudarstvennaya politika regulirovaniya APK [State policy of agribusiness regulation] // monografiya[monograph].– Taraz :Taraz universiteti, 2019. –246 p.
14. Voda dlja ustojchivogo mira. [Текст]: The United Nations World Water Development Report , 2015. PART 1 Italija, 2015. –16-18pp
15. Rogozhina, N.G. Konfliktnyj potencial vodnyh resursov Central'noj Azii [Conflict potential of water resources in Central Asia] // Rossiya i novye gosudarstva Evrazii[Russia and the new Eurasian States] .- 2014. -№1(22).-PP .44-54
16. Rejmere, N.F. Prirodopol'zovanie[Environmental management] / N.F. Rejmere. – Moscow.: Mysl', 1990 . – 637 p.
17. SHvarcev, S.L. Voda kak glavnyj faktor global'noj evolyucii [Water as the main factor of global evolution] // Vestnik Rossijskoj akademii nauk[Bulletin of the Russian Academy of Sciences].- T.83, №2(22).-.-PP.124-131.

МРНТИ 28.23.39

В.С. Йоцов[1]*(orcid - 0000-0002-2860-7918)*
М.С. Туленбаев[2]*(orcid - 0000-0003-0070-4641)*
Н.Н. Керимбаев[3]*(orcid - 0000-0002-3206-0855)*
С.Т. Беглерова[4]*(orcid - 0000-0003-2854-7318)*
С.Ш.Дулатбаева[5]*(orcid - 0000-0001-8155-7334)*

[1,2]Доктор тех.наук, профессор, [3]Доктор пед.наук, доцент,[4]Канд.тех.наук,
[1]University of Library Studies and Information Technologies, Sofia, Bulgaria,
[2,4,5]Таразский региональный университет им. М.Х.Дулати, г. Тараз, Казахстан,
[3]Казахский Национальный университет им.Аль-Фараби, г.Алматы, Казахстан
e-mail: [1]v.jotsov@unibit.bg, [2]mtulenbaev@mail.ru, [3]nurasil@mail.ru,
[4]sbeglerova@mail.ru, [5]dulatbaevasaltanat@mail.ru

ГЛУБОКОЕ МОДЕЛИРОВАНИЕ ДЛЯ ЦЕЛЕЙ КИБЕРБЕЗОПАСНОСТИ: ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ, ПРИМЕНЕНИЯ НАУКИ ДАННЫХ

Аннотация. в статье представлено исследование по сочетанию классических и оригинальных результатов интеллектуальной обработки информации, в переводе: большие данные (big data), data mining, knowledge discovery, advanced analytics, web mining). также показано, как подобные исследования помогают конструировать по-настоящему автономные системы, комбинировать статистический и логический вывод путем эволюционных (постепенных) преобразований. показана практическая невозможность прямого слияния результатов из обоих групп методов,в одну кучу‘. аналитическим путем выявлены основные недостатки, присущие всем методам из цитируемой группы интеллектуального анализа данных. показано, что они преодолимы и даны основные направления и результаты по решению поставленной задачи в исследовательских колективах наших университетов (УНИБИТ, КАЗНУ, Университет Дулати).

Ключевые слова: автономность, интеллектуальный анализ данных, data science (big data, knowledge discovery,data mining, deep data mining, web mining, data analytics, text mining).

Введение. Интеллектуальные методы сегодня находятся в центре современных систем кибербезопасности, в которых только софтверные агенты могут отыскать уязвимости и предотвратить атаки в режиме реального времени и то после анализа больших данных. К сожалению, при таком подходе возникает множество проблем и противоречий, да и требования к такого рода системам возрастают в разы. В данном исследовании показаны синтетические метаметоды, дающие пути для решения накопленных проблем, а также и множество аналитических методов, предназначенные для работы под управлением метаметодов.Чаще всего общая схема управления работает по принципу джунглей (выигрывает наиболее приспособленный, то есть дающий лучшие результаты), т.к. она все еще наиболее проверена и распространена, проста в использовании и программируется быстрее всего. Полученные результаты апробированы в группе систем сетевой безопасности, в системах поиска и предотвращения нарушений, а также для целей создания и применения методов Теории Чисел в крипрографии и стенографии. Основной упор поставлен на использовании методов глубокого моделирования, позволяющих софтверным агентам рассуждать, отыскивать конфликты с базами знаний и со складами данных, осуществлять слияние данных и знаний из разнотипных источников или путем сочетания несовместимых методов.

В информатике издавна существуют проблемы, которые остаются в центре внимания независимо от этапа развития направления или от смены парадигм. Это вопросы конструирования элементов машинного мышления, которые входят в классические направления машинного самообучения(MachineLearning) и обнаружения новых знаний (системы обнаружения скрытых закономерностей, извлечение новых знаний и анализ данных): BigData, KnowledgeDiscoveryandDataMining а также в DataAnalytics, WebMining, TextMining, DeepDataMining). Все они так или иначе ассоциируются с терминами интеллектуального анализа данных [1-6]. По сути, основная цель таких систем состоит в том, чтобы извлечь хотя бы несколько знаний (правил или фактов) из больших массивов неструктурированных данных на основе объединения избранных по определенным критериям данных с использованием разнотипных логических связей. Связывание данных с изпользованием (логических) отношений и есть основа для глубокого моделирования предметной области. Эти знания должны оставаться актуальными для всех типов данных, накопленных до сих пор, независимо от их происхождения, качества, полноты и других параметров. Затем на базе накопленного опыта необходимо приобрести (применить, привнести) определенные мета-знания, то есть знания о том, как наилучшим образом управлять и использовать все наличные знания из базы знаний (БЗ). Не будет преувеличением сказать, что при правильном (эволюционном) использовании знаний и мета-знаний путем применения больших данных может быть сформирована высшая форма информации, которая называется мудростью. В данной статье этот термин мудрость, а также понятие разумной системы используются в точных, специфических терминах в соответствии с базовыми определениями искусственного интеллекта.

Описанные процессы являются как эволюционными, так и итерактивными по природе [2-4]. Для их разработки используют разнообразные подходы, основанные на управлении данными (data-driven approaches), т.е. управлении не по алгоритму, а по ситуации. В настоящее время все большеговоритсяи о непосредственном управлениипосредством знаний (knowledge-driven approaches). Эти две составляющие являются ключевыми для интеллектуальной обработки больших данных, поскольку они позволяют разрабатывать неалгоритмические подходы на основании накопленных данных и с использованием разработанных наборов алгоритмов. За последние 25 лет не было достигнуто значительного прогресса в стандартизации управления данными и их создание является не только научно-прикладным процессом, но в некоторой степени зависит и от искусства и опыта разработчиков. Надеемся, что новые предложения, описанные в следующих разделах статьи, будут способствовать улучшению работы подобных систем.

Переход от огромных множеств неструктурированной информации к получению новых знаний, а затем посредством последовательного накопления опыта (знаний типа мудрости), является сложным и не до конца разработанным процессом с точки зрения человека; все это еще хуже разработано для машин.Для целей данного перехода необходимо использовать методы обработки запросов в базе данных (БД), статистические и логические методы, в том числе искусственный интеллект, логические приложения и машинное самообучение. На вершине этого множества междисциплинарных исследований находятся методы поддержки принятия решений (DecisionSupport). Одним из наиболее сложных вопросов в этой области является то, как объединить статистические и логические результаты в одной системе. Эта проблема подробно рассматривается в следующих разделах данного исследования.

В общем случае данные обрабатываются специальными алгоритмами, по возможности перед этим они проходят и предварительную обработку (препроцесинг) с целью структурирования информации и, при необходимости, ее преобразования в наиболее подходящую форму. Затем к ним применяется интеллектуальная обработка (DataMining): из них (уже на уровне itemsets) выводятся различные законы, производится поиск ассоциаций, анализируются сходство и/или разнотипные взаимосвязи между ними. Наконец, при успешном процессе обработки в БЗ добавляются новые знания и обновленнаяБЗ используется для лучшего, болееэффективного анализа новых входных данных. Так формируется цикл обработки данных.

При работе с большими данными, неминуемо получающимися при работе современных киберзащит, схема аналогична, но обработка намного глубже и, следовательно, более сложна, чем описанный выше пример обнаружения знаний и извлечения данных (knowledge discovery & data mining). Онапо возможности включает в себя работу с агентами, использование онтологий, более сложные процедуры обработки аналитических данных и так далее. Но основные этапы все те же и основной блок здесь – цикл Data Mining. Так в чем же разница? В схеме использования больших данных основное внимание уделено вопросам кибер безопасности, а в классическом варианте (Data Mining) эти вопросы все еще не рассматриваются.

При веб-майнинге (webmining)схема усложнена и используется для большей автономности в системах обработки данных, но общая схема (этапы обработки данных) и здесьаналогична классической.

В аналитике (data analytics) к схеме добавляются новейшие методы статистической обработки данных и глубокого моделирования предметной области.

Несмотря на разнообразие настоящих исследований, все методы из упомянутых областей имеют присущие направлению характерные недостатки. Мы не можем говорить о конвергенции разультатов, полученных с помощью интеллектуальной обработки данных. До сих пор без хорошей подготовки к управлению знаниями и без знания методов искусственного интеллекта нельзя говорить об эффективном управлении интеллектуальными системами. Другими словами, сегодня тот, кто недостаточно подготовлен в интеллектуальном и кибер направлении, не может управлять большими данными и,соответственно, системами кибер безопасности.

Как показано в начале статьи, без Data-Drivenподходовмашина не может быть достаточно автономна. Любой алгоритм кибер защиты в конечном счете может быть разбит нарушителями.В системах должны выявляться и решаться постоянно возникающие смысловые (семантические) конфликты (противоречия). В противном случае, они будут накапливаться, и система будет постепенно деградировать.

Алгоритмическая сложность современных систем обнаружения - по крайней мере экспоненциальна, класс сложности – NPМили NPH, поэтому все еще мало научных разработок в этой области,которые могут найти **широкое** промышленное применение. Сложность разработки комплекса приводит к большей уязвимости к определенным видам атак. Можно сказать, что интеллектуальные системы обработки информации не должны работать без специального комплекса для кибер защиты. Здесь уместно использовать аналогию с авианосцами, кстати это группы с высокой концентрацией интеллектуальных систем. Если авианосец покинет боевую группу, он фактически обречен на гибель.

Следовательно, необходимо разработать сложную систему кибер защиты, т.к. в каждой практически используемой интеллектуальной системе существует большое количество уязвимостей. При этом подсистема для кибер безопасности должна обладать определенной автономией, иначе человек-нарушитель или продвинутая программа легко ее ‚раскусят‘. В данной подсистеме должны использоваться софтуерные агенты с целью обработки больших данных – нечеловеческое это дело перебирать и анализировать миллиарды и триллионы данных и знаний. Выходит, что для качественной работы с большими данными необходимо постоянно совершенствовать как интеллектуальную так и кибер- подсистемы.

И, наконец, необходимо определиться какое количество данных является действительно большими данными. Неспроста в разных источниках упоминаются разные цифры, а большинство исследователей просто избегают дикуссии в этом направлении. Прежде всего, количество определяется в зависимости от поставленных перед системой задач. Например, если поставлена задача скопировать человека на атомарном уровне, то все современные большие БД являются все еще ничтожно малыми для решения проблем. Во вторых, время обработки информации - также существенный фактор. Если у нас есть система с миллионом данных для отыскания орудует ли вор-профессионал в супермаркетах города, то это - большие данные. Если

необходимо просто перечислить миллион телефонов из одного города, то это – не особенно большие данные. Большие БД могут работать и с небольшими массивами даных, это нормальная, хотя и достаточно затратная практика.

1. **Объединение логического и статистического анализа в одной системе.** В области интеллектуальной обработки больших данных существует более 11 наборов стандартов для обработки аудио-, мультимедийных и других типов сигналов или других типов сенсорной информации (данных). Тем не менее, большинство исследований в этом направлении не стандартизированы или стандарты больше воспринимаются в качестве рекомендаций или передового опыта. В этой ситуации мы предложили новые и нестандартные решения в этой области.

Использование различных статистических методов является широкой практикой как в преподавании, так и в исследованиях в области науки данных для кибер безопасности. В последнее время все больше и больше статистических приложений называют интеллектуальными в основном из-за их растущей популярности и хорошего финансирования. С другой стороны, использование каких-либо статистических или других алгоритмических средств не делает системы интеллектуальными. Чаще всего статистические разультаты должны обрабатываться и на логическом уровне - теория доказательств (EvidenceTheory) используется, например, в следующих исследованиях [7-9]. Но всего этого недостаточно для принятия необходимых решений во многих случаях. Теория позволяет нам правильно интерпретировать полученные цифровые результаты, но ничто не говорит о том, являются ли полученные результаты достаточно надежными. Кроме того, теория подтверждения широко использует эвристические оценки типа функций убеждения, функций сомнения (belief functions, doubt functions), что также ставит под вопрос результаты обработки.

Вероятностные методы все чаще используются из-за их высокой эффективности. Например, из-за относительно низкой алгоритмической сложности приложений. Однако даже при вероятности из доверительного интервала 1 полученная гипотеза с такой высокой вероятностью(например 0.9999) все еще не может считаться доказанной (вспомним пример черного лебедя из экономики). Именно поэтому специалисты разделяют платформы интеллектуального анализа данных (DataScience) на две независимые части - вероятностную и логическую. В этой ситуации правильная работа с интеллектуальными платформами обработки больших данных - это вопрос глубокого познания теории интеллектуальных систем. По той же причине мы пока не можем говорить о качественном автономном управлении интеллектуальными платформами. Нельзя сваливать в одну кучу результаты статистической и логической обработки больших данных. Необходимо для слияния (fusion) такого рода результатов разрабатывать и/или применять специальные управляемые данными эволюционные методы, как показано в следующем разделе.

2. **Описание исследований.** От достижений, полученных при помощи вероятностных методов, сегодня невозможно отказаться. Сегодня они достаточно эффективны, но чаще всего это достигается за счет инкапсуляции информации. С другой стороны, логические методы чаще всего не могут работать с той же минимальной алгоритмической сложностью, однако они не инкапсулируют используемую информацию, а успешно служат для извлечения новых знаний. Слияние данных двух групп принципиально разнотипных результатов не может быть достигнуто механическим путем

[10-13]. Для этой цели должны использоваться различные эволюционные процессы, в том числе применение элементов примитивного (в настоящее время) машинного мышления, которое чаще всего определяется как последовательная эволюционная обработка данных и знаний.

Предлагается использовать статистические методы в ситуациях, пока не накоплено достаточно много знаний по предмету (в среде с недостаточными знаниями: knowledge-poorenvironments). В этом случае основная причина использования вероятностных методов заключается в следующем: лучше слабая и ненадежная информация, чем полное ее отсутствие. Поэтому при отсутствии накопленного опыта в системе (например, у автономных агентов) вероятностные или аналогичные им по природе оценки (нечеткие и т.д.) будут используются в принудительном порядке. Однако с накоплением достаточного опыта в системе (существуют методы, разработанные для таких оценок), оценки вероятности постепенно **отменяются** (важность их использования падает, где и когда возможно, они исключаются из среды управления более универсальными и гибкими логическими методами). Независимо от способов реализации, описанные процессы носят эволюционный характер.

При объединении результатов, предложенных двух типов методов используется принцип (survival of the fittest)'выживание наиболее приспособленных. Таким образом, решения, которые изначально оцениваются как слабые или противоречивые, могут в будущем продолжать развиваться и даже могут дать впоследствии самые сильные результаты. Они не отсеиваются, как в большинстве современных методов, потому что любая оценка может быть мгновенной и/или субъективной: в определенных ситуациях ей надо дать возможность для развития (переоценки). Единственная бесспорная оценка - это та оценка, что доказана. Такое отсутствует во многих из вышеупомянутых методов и поэтому недоразвитое потенциально сильное решение можно спутать со слабым и непредсказуемым.

Вот формальные описания процедур отмены(анулирования) решений (DefeasibleReasoning). Используются следующие правила (Horn clauses):

$$B \leftarrow \bigwedge_{i \in I} A_i . \tag{1}$$

Логический ввод изменяется, когда к правилу прилагается исключение, анулирующее атом (конъюнкт) A_k, которое записывается следующим образом: E(C, Ak). Если C (причина для анулирования) истинно и если Ak - ложь (см. ниже), значение правила для доказательства **B** изменяется (по исключению). В порядке исключения предметная область была пополнена новыми знаниями, что, в свою очередь, привело к изменению значения правила (см. аналогичные исследования [14,15]). В расширенной теории анулирующего заключения используются следующие и другие формальные описания, наиболее часто используемое из которых состоит в следующем:

$$\frac{B \leftarrow \bigwedge_{i=1}^{z} A_i, C, E(C, A_k), \neg A_k \leftarrow C}{B \leftarrow A_1 \wedge A_2 \wedge \ldots A_{k-1} \wedge \neg A_k \wedge \ldots A_z} \tag{2}$$

Другими словами, описанные анулирующие приложения могут исключать или сводить к нулю любую информацию вероятностного или логического характера (сюда включаем и невидимые логические связи между частями-атомами знаний). С другой стороны, они не всегда исключают ее при оценках с низким значением - как некоторая информация была отменена,

так других случаях ее важность может восстановиться и даже увеличиться. Таким образом, управление предлагаемыми системами периодически берет на себя как статистическая, так и логическая подсистемы, но по истечении определенного периода времени статистика постепенно теряет свою доказательную ценность[16].

Другой способ реализации эволюционных стратегий состоит в использовании аппарата эволюционных игр [17,18].

В данном разделе представлен метод выявления и разрешения противоречий путем сравнения знаний с заранее введенными схемами противоречий. Если в модели нет явно поставленныхусловий для существования противоречия, она называется синтаксической моделью [19]. В противном случае модель является семантической. Представленные модели можно рассматривать как вид онтологии для обнаружения противоречий в БЗ. После обнаружения противоречий они разрешаются в соответствии с основными направлениями, описанными в книге [19].

Использование разрешения противоречий и анулирования определенных знаний все еще недостаточно для разработки и дальнейшего развития систем и платформ обработки больших данных [20,21]. Метод для определения связи между известными данными и знаниями с неизвестными способами решения одной или нескольких назначенных задач описан ниже. С помощью Puzzle Methods, разработанных в Unibit, цели достигаются с помощью применения различных типов ограничений. Поставленные цели также могут быть осуществены с помощью системы линейных или нелинейных ограничений, а затем путем последовательного рассмотрения решений в полученной узко ограниченной области.

Рисунок 1иллюстрирует пример использования трех элементов K = {A, C, E} иL = {B, D}. В примере используются и классические ограничения V_u.

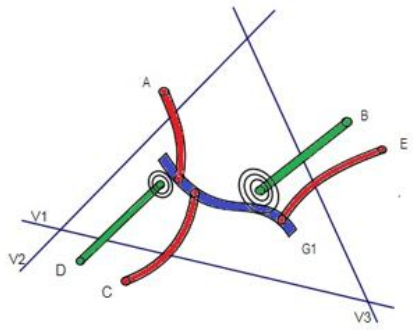

Рис. 1. Разнотипные виды ограничений для решения поставленной цели

В результате использования всех трех V_u формируется ограниченная область, в которой легче искать требуемое решение (Цель G [Goal] для выполнения задачи ConstraintSatisfaction). К сожалению, трудно получить такое множество ограничений V_u, которое приведет к требуемому результату. Пример иллюстрирует преимущества введения элементов логически-базированных множествL и K, гдеограничения из K названы ограничениями кроссворда – пресечение A, C или E с областью искомой целиG_1 дают частичное решение, которое может быть связано с другими знаниями для выявления цели. Они хорошо сочетаются с классическими пространственными ограничениями(линейными, нелинейными, бинарными и др.)V_u. В то же время, элементы множества K не дают части искомого

решения, а указывают, что оно где-то рядом. В данном случае удобно использовать аппарат нечетких логик. В целом, местоположение G_1 не помогает в определении решения подзадач, и введение D, E…V_uприводит к быстрому уменьшению количества рассматриваемых альтернатив. Здесь показаны и новые (логические) типы ограничений, с которыми мы имели дело - те, которые логически связывают цель с накопленными ограничениями, и те, которые показывают, насколько близка желаемая цель G_1. Еще один тип неклассических ограничений – указвающие ограничения: в каком направлении искать. Можно сказать, что частным случаем данного типа неклассичеких ограничений является целевая или фитнес функция.

В результате применения глубокого моделирования, например, с использованием представленных трех групп ограничений, можно получить удовлетворительные решения по применению технологии больших данных для целей кибер-безопасности. Конечно, можно не использовать все перечисленные средства в одном комплексе, однако в связи с быстрым развитием технологий применение недоработанных или однобоких решений не рекомендуется.

3. **Введение в реализации.** Используемые в данном исследовании экспериментальные приложения можно разделить на две группы. Первая группа характерна для большинства информационных приложений, и эксперименты описаны в ряде наших публикаций в 2013-2019 гг. В [16] обсуждается, как использовать PuzzleMethods для улучшения производительности SASEnterprizeMiner 1.12 – данная платформа является одной из лучших в среде DataMiners. Один из основных экранов продукта представлен на рисунке 2.

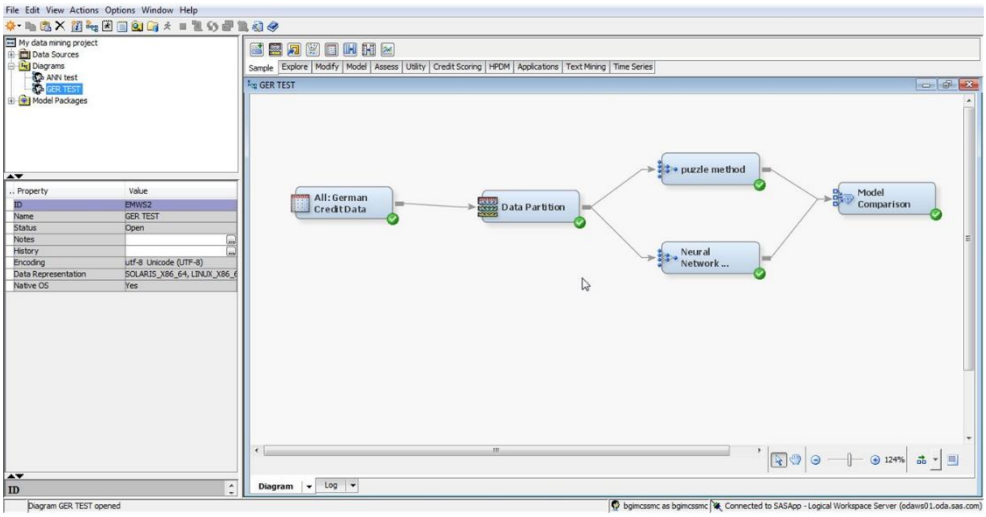

Рис. 2.Экран работы ссистемой SAS 1.12

Ввиду универсальности предложенных эволюционных решений и вопреки известным парадигмам, областью применения некоторых наших экспериментов были такие абстрактные области, как Теория Чисел с приложениями для криптографических целей. Исследование началось с построения эвристических догадок, гипотез, которые прошли описанные эволюционные процессы, которые привели к улучшению и отсеиванию недостоверной информации. Наконец, появился ряд интересных

свидетельств, и в сопутствующем аспекте - разработка новых, неизвестных шифров и мотодов как для интеллктуальных так и для кибер-применений.

Заключение. Показаны основные направления и практические вопросы применения интеллектуальной обработки больших данных, а также и проблемы, связанные с этим направлением: разрешение противоречий, связывание неизвестной информации с известными (накопленными) знаниями с целью решения поставленных задач, объединение результатов статистических и логических приложений в одну систему и использование логически-базированной системы ограничений для поска решений. Показано, что управление этими процессами должно носить эволюционный характер и неминуемо приведет к появлению ряда интересных прикладных и абстрактных новых методов и широкого круга систем в области кибер безопасности.

Список литературы

1. Duch, V. (2006) Computational Creativity. // Proc. Int. Joint. Conf. on Neural Networks, – Vancouver, BC, Canada, – July 16-21, – P. 435-442.
2. Towns, J. et. al. (2014) XCEDE: Accelerating Scientific Discovery. // J. Computing in Science and Engineering, – September/October 2014, –P. 62-74.
3. Otero, C., Peter, A. (2015) Research Directions for Engineering Big Data Analytics Software. // IEEE J. Intelligent Systems, –January/February 2015, – P. 13-19.
4. El-Gayar, O., Leung, P., and Scharl, A. (2013) Introduction to analytics, Informatics and Decision Support for Sustainability Minitrack. // 46th Hawaii Intl. Conf. on System Sci., – P. 914-919.
5. Aytkhozhaeva E., Seilova N. (2017) Information security of the electronic society and the internet of things. / J. News of the National Academy of Sciences of the Republic of Kazakhstan, – Series of Geology and Technical Sciences, – Volume 6, Number 426, – P. 264 – 272
6. Samigulina G., Nyusupov A., Shayakhmetova A. (2018) Analytical review of software for multi-agent systems and their applications. / J. News of the National Academy of Sciences of the Republic of Kazakhstan, – Series of Geology and Technical Sciences, – Volume 3, Number 429, – P. 173 – 181
7. Han, D., et al. (2011) New Dissimilarity Measures in Evidence Theory. // Proc. of the 14th IEEE International Conference on Information Fusion, – 5-8 July 2011, – Chicago, IL, –P.1-7.
8. Yong, W., et al. (2011) Research on Evaluation Method of Power Quality for Wind Plant Based on Probability Theory and Evidence Theory. // International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), – December 16-18, Changchun, China, – P. 1003-1006.
9. Dong G., Kuang, G. (2015) Target Recognition via Information Aggregation Through Dempster–Shafer's Evidence Theory. // IEEE J. Geoscience And Remote Sensing Letters, – Vol. 12, No. 6, – June 2015, – P. 1247-1251.
10. Назаров А., Комаров А. (2013) Интеллектуальная система анализа кибербезопасности в пространстве на web-технологиях. / J. T-Comm, #10, – 2013, – P. 81-84
11. Davidson-Pilon C. (2019) Bayesian Methods for Hackers and DataOrigami. – Back to school for Food Science. – @Shopify. Waterloo, Canada. – https://github.com/CamDavidsonPilon to date
12. Emmert-Streib F., Dehmer M. (2019) J. Mach. Learn. Knowl. Extr. 2019,– 1,– P. 945–961
13. Domingos, P. et. al. (2019) Unifying Logical and Statistical AI. – https://homes.cs.washington.edu/~pedrod/papers/aaai06c.pdf to date
14. Bryant D., Krause P. (2019) A review of current defeasible reasoning implementations. – https://www.semanticscholar.org/paper/A-review-of-current-

defeasible-reasoning-Bryant-Krause/532868041cefcd5211ffe-4021823a43c1a6b7cee to date
15. Casini G. et. al (2019) Towards Practical Defeasible Reasoning for Description Logics. – http://ceur-ws.org/Vol-1014/paper_17.pdf to date
16. Jotsov, V. (2016) New Proposals for Knowledge Driven and Data Driven Applications in Security Systems. // Innovative Issues in Intelligent Systems. V. Sgurev, R. Yager, J. Kacprzyk, V. Jotsov (Eds.) – Studies in Computational Intelligence, – vol. 623, Springer, – Berlin Heidelberg, –P. 231-294 (ISSN: 1860-949X).
17. Karev G. (2018) Evolutionary games: natural selection of strategies. – https://arxiv.org/ftp/arxiv/papers/1802/1802.07190.pdf to date
18. Newton J. Evolutionary Game Theory: A Renaissance. – http://res.mdpi.com to date
19. Йоцов, В. (2014) Изкуствен интелект и експертни системи. – София: За буквите – О писменехъ, – 2014, – 236 с.
20. Big Data, Big Promises: The Next Generation of Conflict Forecasting (2018) Friday, – 26 January, 2018 – https://www.ethz.ch › ethz › dual to date
21. Benefits and risks of Big Data Analytics in fragile and conflict affected states (2019) – https://assets.publishing.service.gov.uk/media/5d1c7da440f0b609cfd974a1/605_Benefits_and_Risks_of_Big_Data_Analytics_in_Fragile_and_Conflict_Affected_States__FCAS_.pdf to date

Материал поступил в редакцию 09.06.21.

В.С. Йоцов[1], М.С. Туленбаев[2], Н.Н. Керимбаев[3], С.Т. Беглерова[4], С.Ш.Дулатбаева[5]

[1] Кітапхана ісі және ақпараттық технологиялар университеті, София, Болгария,
[2,4] М.Х.Дулати атындағы Тараз өңірлік университеті, Тараз, Қазақстан,
[3] Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

КИБЕРЛІК ҚАУІПСІЗДІК ҮШІН ТЕРЕҢ ҮЛГІ: МӘЛІМЕТТЕР ҒЫЛЫМЫН ҚОЛДАНУҒА ҚИЫНДЫҚТАР МЕН ПЕРСПЕКТИВАЛАР

Аннотация. Мақалада ақпаратты интеллектуалды өңдеудің классикалық және түпнұсқа нәтижелерін үйлестіру туралы зерттеу ұсынылған, аудармада: үлкен деректер, деректерді жинау, білімді ашу, дамыған аналитика, веб-тау-кен өндірісі). Сонымен қатар, мұндай зерттеулер автономды жүйелерді құруға, эволюциялық (біртіндеп) қайта құрулар арқылы статистикалық және логикалық қорытындыларды біріктіруге қалай көмектесетінін көрсетеді. Екі топтағы әдістердің нәтижелерін тікелей бір үйіндіге біріктіру мүмкін емес екендігі көрсетілген. Аналитикалық тұрғыдан деректерді өндірудің келтірілген тобынан барлық әдістерге тән негізгі кемшіліктер анықталды. олардың жоғары екендігі көрсетілген және біздің университеттердің ғылыми топтарындағы проблеманы шешу үшін негізгі бағыттар мен нәтижелер берілген (ҚазҰУ, ТарӨУ).

Тірек сөздер: автономия, деректерді өндіру, деректер ғылымы (үлкен деректер, білімді ашу, деректерді өндіру, терең деректерді өндіру, веб-тау-кен жұмыстары, деректерді талдау, мәтіндерді өндіру).

V.S. Yotsov[1], M.S.Tulenbaev[2], N.N.Kerimbaev[3], S.T.Beglerova[4], S.Sh.Dulatbaeva[5]

[1]University of Library Studies and Information Technologies, Sofia, Bulgaria,
[2.4]Taraz Regional University named afterM.Kh. Dulati, Taraz, Kazakhstan,
[3]Kazakh National University named after Al-Farabi, Almaty, Kazakhstan

DEEP MODELING FOR CYBER SECURITY: CHALLENGES AND PROSPECTS FOR THE APPLICATION OF DATA SCIENCE

Abstract. The paper considers the importance of intelligent cyber security applications and contemporary problems and features of data science (big data, data mining, knowledge discovery, advanced analytics, web mining, text mining), and other intelligent data processing systems. The novel results in all of the quoted fields rely on realizations of autonomous systems, applications of data-driven methods, fusion of statistical and logical results, deep processing of accumulated knowledge, etc. The intelligent technologies advance the efficiency of statistical applications in one evolutionary process. Novel results had been elaborated for evolutionary methods supporting the fusion process in the logical-and-statistical complex. It is shown how data science methods can improve the quality of many contemporary applications in the cyber security fields. All the quoted advantages may be successfully combined with classical and other known applications.It is demonstrated that without the described research innovations the main cyber security shortcomings from the considered contemporary intelligent systems couldn't be completely resolved.

Keywords: DataSceince, Cyber Security, Autonomous System, Software Agent, Intelligent Systems