

## МЕТРИКА ДЛЯ АНАЛИЗА BIG DATA

*ХАХАНОВ В.И., МИЩЕНКО А.С.,  
ОБРИЗАН В.И., TAMER BANI AMER*

Предлагается инфраструктура обеспечения параллельного анализа big data для поиска, распознавания и принятия решений на основе использования предложенной метрики киберпространства. Используется метрика анализа киберпространства (big data), которая характеризуется наличием единственной логической хог-операции для определения киберрасстояния путем циклического замыкания не менее одного объекта, что дает возможность на порядок повысить быстродействие анализа данных.

### 1. Введение

Киберфизическая система призвана сделать активной концепцию big data, рассматривая большие данные во взаимодействии с киберсистемами (облаками) управления, ориентированными на поиск, распознавание и принятие решений. Структурное содержание CPS – совокупность коммуникационно связанных реальных и виртуальных компонентов с выраженными функциями адекватного физического цифрового мониторинга и оптимального облачного компьютерного киберуправления для обеспечения качества жизни, продукции, процессов или сервисов в заданных условиях ограничений на время и ресурсы. CPS включает компоненты: Cyber Control, Internet of Things или Cloud, Security, Intelligence, Big Data and Services, Digital Monitoring, Cyber Managing, Physical Smart Everything, Nature, Social, and Tech World. Регуляторные отношения (Relationship) между компонентами CPS формируются законами, уставами предприятий и организаций, морально-этическими правилами поведения внутри социальной группы. Направление движения RoadMap – Harmony of Human, Nature and Tech киберфизической системы человечества можно определить как достижение такого интегрального уровня развития киберфизических компонентов, который обеспечит гармонию жизни человека с природой и техникой (созданным миром – Created World).

Big Data – технологическая культура киберпространства, направленная на формирование динамически развивающейся инфраструктуры киберфизической экосистемы планеты путем семантической структуризации больших потоков (объемов) гетерогенных данных на основе использования интеллектуальных быстродействующих специализированных облачных фильтров параллельного мониторинга и метрического анализа извлекаемой информации для online управления физическими и виртуальными процессами.

Big data driven киберфизическая система управления физическими процессами имеет инновационные отличия от существующих решений, в частности, использование нецифровой булеан-метрики для структурного и скалярного оценивания процессов и явлений.

Цель big data – выращивание в киберпространстве культурного слоя инфраструктуры метрико-семантически упорядоченной легко доступной полезной reusable информации за счет разработки виртуальных облачных сервисов, выполняющих роль быстродействующих интеллектуальных фильтров в процессах поиска, распознавания и принятия решений.

Отсюда одна из задач технологической культуры big data – создание метрик измерения (дифференцирования) процессов и явлений в киберпространстве для построения быстродействующих семантических и оценивающих фильтров поиска полезной информации.

### 2. Метрика для анализа big data

Идея, способствующая повышению быстродействия анализа big data – убрать из виртуального компьютера «тяжеловесную» арифметику и трансформировать освободившиеся ресурсы для создания параллельного виртуального мультипроцессора векторно-логических вычислений в целях поиска, распознавания и принятия решений с помощью примитивных операций: and, or, not, хог. Специализация компьютерного изделия, ориентированная на использование только логических операций, дает возможность существенно ( $\times 100$ ) повысить быстродействие решения неарифметических задач. Исключение арифметических операций, использование параллелизма алгебры векторной логики, мультипроцессорность архитектуры создают эффективную инфраструктуру, которая объединяет математическую и технологическую культуру для решения прикладных задач анализа киберпространства. Рыночная привлекательность логического виртуального мультипроцессора (Logical Virtual MultiProcessor – LVMP) определяется сотнями старых и новых логических по своей природе задач, которые в настоящее время решаются не всегда неэффективно на избыточных универсальных компьютерах с мощными арифметическими процессорами.

Цель исследования – существенное повышение быстродействия процедур поиска, распознавания и принятия решений путем мультипроцессорной и параллельной реализации ассоциативно-логических векторных операций для анализа больших данных в векторном логическом пространстве без использования арифметических операций.

Для достижения поставленной цели необходимо разработать неарифметическую метрику оценивания векторно-логических решений в кибернетическом big data пространстве.

Объектом исследования является инфраструктура процессов поиска, распознавания и принятия решений в big data пространстве на основе использования алгебры векторной логики, квантовых структур данных, вычислительной архитектуры анализа ассоциативно-логических структур данных и неарифметического интегрального критерия качества.

В процессе исследований использованы источники научно-технической информации: ассоциативно-ло-

гические структуры данных для решения информационных задач [2]; аппаратная платформа векторно-логического анализа информации; модели и методы дискретного анализа и синтеза объектов киберпространства [3], а также кубитное представление дискретных объектов [1].

Дискретное векторно-логическое пространство (киберпространство) – совокупность взаимодействующих по соответствующей метрике информационных процессов и явлений, описываемых векторами логических переменных и использующих в качестве носителя компьютерные системы и сети.

Метрика – способ измерения расстояния в киберпространстве между компонентами процессов или явлений, описанных векторами логических переменных, задаваемых в булевом или булеановом (многозначном) алфавите. Расстояние в киберпространстве – это хог-отношение (симметрической разности) между отличным от нуля числом векторов, обозначающих компоненты процесса или явления, что отличает его от кодового расстояния по Хэммингу. Расстояние, производная (булева), степень изменения, различия или близости есть изоморфные понятия, связанные с определением отношения двух компонентов процесса или явления. Понятие близости (расстояния) компонентов в киберпространстве есть мера их различия. Процедуры сравнения, измерения, оценивания, распознавания, тестирования, диагностирования, идентификации есть способ определения отношения при наличии не менее, чем двух объектов.

Компонент пространства представлен  $k$ -мерным (двоичным) вектором  $a = (a_1, \dots, a_j, \dots, a_k)$ ,  $a_j \in \{0, 1\}$ , где каждая его координата определена в двоичном алфавите, 0 – «ложь», 1 – «истина». Нуль-вектор есть  $k$ -мерный кортеж, все координаты которого равны нулю:  $a_j = 0, j = \overline{1, k}$ .

Метрика  $\beta$  кибернетического пространства определяется единственным равенством, которое формирует нуль-вектор для хог-суммы расстояний  $d_i$  между ненулевым и конечным числом точек (объектов), замкнутых в цикл

$$\beta = \bigoplus_{i=1}^n d_i = 0, \quad (1)$$

где  $n$  – количество (целое число) расстояний между компонентами (векторами) пространства, составляющими цикл  $D = (d_1, \dots, d_i, \dots, d_n)$ ,  $d_i$  – вектор расстояния, соответствующий ребру цикла, соединяющему два компонента (вектора)  $a, b$  пространства, который далее обозначается без индекса как  $d(a, b)$ . Расстояние между двумя объектами (векторами)  $a$  и  $b$  есть производный вектор:  $d(a, b) = (a_j \oplus b_j)_1^k$ . Векторному значению расстояния соответствует норма – скалярное расстояние по Хэммингу между

двумя векторами – как число единиц вектора  $d(a, b)$ . Иначе: метрика  $\beta$  векторного логического двоичного пространства есть равная нуль-вектору хог-сумма расстояний между конечным числом точек (вершин) графа, образующих цикл. Сумма  $n$ -мерных двоичных векторов, задающих координаты точек циклической фигуры, равна нуль-вектору. Свертка пространства в нуль-вектор представляет интерес для многих практических задач, включая: диагностирование и исправление ошибок при передаче информации по каналам связи; поиск дефектов в цифровых изделиях на основе двузначных и многозначных таблиц неисправностей. Кроме того, на основе введенной метрики можно дать более формальное определение киберпространства, которое является векторно-логическим, нормируемым  $\beta$ -метрикой, где хог-сумма расстояний между конечным числом точек цикла равна нуль-вектору. Определение метрики ставит во главу угла не элементы множества, а отношения, что позволяет сократить систему аксиом (тождественности, симметрии и транзитивного треугольного замыкания) с трех до одной и распространить ее действие на сколь угодно сложные конструкции  $n$ -мерного логического пространства. Классическое задание метрики для определения взаимодействия одной, двух и трех точек в векторном логическом пространстве является частным случаем  $\beta$ -метрики при  $i=1, 2, 3$  соответственно:

$$M = \begin{cases} d_1 = 0 \leftrightarrow a = b; \\ d_1 \oplus d_2 = 0 \leftrightarrow d(a, b) = d(b, a); \\ d_1 \oplus d_2 \oplus d_3 = 0 \leftrightarrow d(a, b) \oplus d(b, c) = d(a, c). \end{cases} \quad (2)$$

Векторно-логический транзитивный треугольник (2) имеет полную аналогию численному измерению расстояния в метрическом  $M$ -пространстве, которое задается системой аксиом, определяющей взаимодействие одной, двух и трех точек в любом пространстве:

$$M = \begin{cases} d(a, b) = 0 \leftrightarrow a = b; \\ d(a, b) = d(b, a); \\ d(a, b) \oplus d(b, c) \geq d(a, c). \end{cases} \quad (3)$$

Специфика аксиомы треугольника (3) метрического пространства заключается в численном (скалярном) сравнении расстояний трех объектов, когда интервальная неопределенность ответа – две стороны треугольника могут быть больше либо равны третьей – малоприспособна для определения точной длины последней стороны. Векторно-логическое пространство устраняет данный недостаток, полностью исключает степень неопределенности в бинарном отношении детерминированных состояний процессов или явлений. В этом случае численная неопределенность третьей стороны треугольника в векторном логическом пространстве приобретает форму точного двоичного вектора, который характеризует расстояние между двумя объектами и вычисляется на основе знания расстояний двух других сторон треугольника:

$$d(a, b) \oplus d(b, c) = d(a, c) \rightarrow d(a, b) \oplus d(b, c) \oplus d(a, c) = 0.$$

*Пример 1.* Имеется пять точек в векторном пространстве: (000111, 111000, 101010, 010101, 110011). Замыкание этих точек в цикл дает следующие стороны-расстояния в пятиугольнике: (111111, 010010, 111111, 100110, 110100).

Покоординатное сложение всех векторов дает результат (000000). Практическая значимость данного факта заключается в возможности восстановления любого расстояния в замкнутом цикле, если известны (n-1) сторон фигуры. Для треугольника это означает восстановление третьей стороны по известным двум. Если же создать из треугольников замкнутое логическое пространство, то можно сэкономить 66% от объема данных, который формирует все расстояния в логическом пространстве.

Метрика  $\beta$  кибернетического многозначного векторно-логического пространства есть вектор, равный значению  $\emptyset$  по всем координатам, полученный путем применения симметрической разности расстояний между конечным числом точек, образующих цикл:

$$\beta = \Delta_{i=1}^n d_i = \emptyset, \quad (4)$$

где каждая координата вектора, соответствующего объекту, определена в алфавите, составляющем булеан на универсуме примитивов мощностью  $p$ :  $a = (\alpha_1, \dots, \alpha_j, \dots, \alpha_k), m = 2^p$ .

Равенство пустому вектору симметрической разности покоординатного теоретико-множественного взаимодействия (4) подчеркивает равнозначность компонентов (расстояний), участвующих в формировании уравнения, где единственная координатная операция  $d_{i,j} \Delta d_{i+1,j}$ , используемая, например, в четырехзначной модели Кантора  $A = \{0, 1, x, \emptyset\}, x = \{0, 1\}$ , определяется соответствующей таблицей:

$\Delta$	0	1	x	$\emptyset$
0	x	1	0	
1	x	$\emptyset$	1	
x	1	0	x	
$\emptyset$	0	1	x	

$\cap$	0	1	x	$\emptyset$
0	0	$\emptyset$	0	
1	$\emptyset$	1	1	
x	0	1	x	
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	

$\cup$	0	1	x	$\emptyset$
0	0	x	x	0
1	x	1	x	1
x	x	x	x	x
$\emptyset$	0	1	x	$\emptyset$

a	0	1	x	$\emptyset$
$\tilde{a}$	1	0	$\emptyset$	x

 $(5)$

Здесь также приведены таблицы истинности для других базовых теоретико-множественных операций (пересечение, объединение, дополнение), далее используемых по тексту. Число примитивных символов, образующих замкнутый относительно теоретико-множественных координатных операций алфавит, может быть увеличено. При этом мощность алфавита (булеана) определяется выражением  $m = 2^p$ , где  $p$  – число примитивов. Для практического использования введенной метрики киберпространства далее предлагается доказательный переход от численной характеристики бинарного отношения объектов, объединяющей

три скалярные оценки их взаимодействия, к чисто векторно-логическому критерию качества отношения двух объектов.

*Пример 2.* Замыкание любых сколь угодно сложных конструкций или фигур (линии, плоскости, кубы произвольной мерности) в киберпространстве. Пусть имеются векторы, заданные в алфавите Кантора  $\{0, 1, X\}$ : {010XXX10, 10X00XX1, XX0011XX, 1100XX00}. Расстояния между ними определяются на основе использования  $\Delta$  – операции симметрической разности (XOR-функция, заданная в алфавите Кантора). Кубы, задающие расстояния между исходными векторами, вычисленные путем применения введенной координатной операции, имеют следующий вид:

	0	1	0	X	X	X	1	0
$\Delta$	1	0	X	0	0	X	X	1
=	X	X	1	1	1	$\emptyset$	0	X

	1	0	X	0	0	X	X	1
$\Delta$	X	X	0	0	1	1	X	X
=	0	1	1	$\emptyset$	X	0	$\emptyset$	0

	X	X	0	0	1	1	X	X
$\Delta$	1	1	0	0	X	X	0	0
=	0	0	$\emptyset$	$\emptyset$	0	0	1	1

	1	1	0	0	X	X	0	0
$\Delta$	0	1	0	X	X	X	1	0
=	X	$\emptyset$	$\emptyset$	1	$\emptyset$	$\emptyset$	X	$\emptyset$

Естественно, что XOR-сумма всех расстояний между фигурами или объектами, образующими замкнутый цикл в киберпространстве, равна пустому вектору:

$\Delta$	X	X	1	1	1	$\emptyset$	0	X
	0	1	1	$\emptyset$	X	0	$\emptyset$	0
	0	0	$\emptyset$	$\emptyset$	0	0	1	1
	X	$\emptyset$	$\emptyset$	1	$\emptyset$	$\emptyset$	X	$\emptyset$
=	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

*Пример 3.* Стрельба по мишени может быть интерпретирована взаимодействием двух матриц, в результате чего появляется третья, идентифицирующая успешность выполненной операции, которая имеет структурный и скалярный эквиваленты, необходимые для online коррекции действий киберсистемы управления, направленных на достижение идеального результата при использовании минимальных временных и материальных затрат:

X	X	X	X	X
X	1	0	1	X
X	0	0	0	X
X	1	0	1	X
X	X	X	X	X

$\emptyset$	$\emptyset$	X	X	X
$\emptyset$	$\emptyset$	0	1	X
$\emptyset$	$\emptyset$	0	0	X
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

X	X	$\emptyset$	$\emptyset$	$\emptyset$
X	1	$\emptyset$	$\emptyset$	$\emptyset$
X	0	$\emptyset$	$\emptyset$	$\emptyset$
X	1	0	1	X
X	X	X	X	X

В данном примере структура результирующей матрицы показывает области, подлежащие уничтожению, а также какого типа снаряды должны быть использованы при стрельбе. Кроме того, нетрудно посчитать эффективность произведенного выстрела, как отношение пора-

женной области к совокупному пространству с учетом значимости (мощности подмножества примитивных символов) каждой координаты матрицы для выбранной цели:

$$\Omega = \frac{\sum_{i=1,n} \sum_{j=1,m} M_{ij}^p \times k_i}{\sum_{i=1,n} M_{ij} \times k_i} = \frac{14}{41} = 0,341.$$

### 3. Выводы

Существующие программные продукты и публикации практически не предлагают ассоциативно-логических технологий поиска, распознавания и принятия решений в дискретном информационном пространстве, состоящем из big data. Практически все они используют универсальную систему команд современного дорогостоящего процессора с математическим сопроцессором. В то же время, аппаратные специализированные средства логического анализа, являющиеся их прототипами [15], как правило, ориентированы на побитовую или не векторную обработку информации.

Фактическая реализация подхода основана на предложении инновационных моделей и методов, использующих идею векторно-логической метрики киберпространства.

Метрика анализа киберпространства (big data), которая характеризуется использованием единственной логической хог-операции для определения кибер-расстояния путем циклического замыкания не менее одного объекта, что дает возможность на порядок повысить быстродействие анализа big data и подсчет структурных критериев качества взаимодействия информационных объектов на основе использования векторных логических операций для точного поиска, распознавания образов и принятия решений.

Будущие исследования направлены на проектирование big data driven cyber physical systems, которые ориентированы на постоянную метрико-семантическую реструктуризацию киберпространства в целях удобного извлечения знаний, а также на преобразование социальных отношений неприродного мира путем передачи управления от человека к облачным сервисам.

**Литература:** 1. *Бондаренко М.Ф., Хаханов В.И., Литвинова Е.И.* Структура логического ассоциативного мультипроцессора // Автоматика и телемеханика. 2012. № 10. С. 71-92. 2. *Ariane Hellinger, Ariane Hellinger, Heinrich Seeger.* Cyber-Physical Systems. Driving force for innovation in mobility, health, energy and production. Acatech. National Academy of Science and Engineering. 2011. 48p. 3. *Hahanov V.I., Litvinova E.I., Chumachenko S.V.* et al. Qubit Model for solving the coverage problem // Proc. of IEEE East-West Design and Test Symposium. Kharkov. 14-17 September, 2012. P.142-144.

Поступила в редколлегию 11.05.2014

**Рецензент:** д-р техн. наук, проф. Чумаченко С.В.

**Хаханов Владимир Иванович**, декан факультета КИУ, д-р техн. наук, профессор кафедры АПВТ ХНУРЭ. Научные интересы: проектирование и тестирование вычислительных систем, сетей и программных продуктов. Увлечения: баскетбол, футбол, теннис, горные лыжи. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. +380 57 70-21-326, E-mail: hahanov@kture.kharkov.ua.

**Обризан Владимир Игоревич**, старший преподаватель кафедры АПВТ ХНУРЭ. Научные интересы: облачные технологии, программирование мобильных платформ. Увлечения: путешествия. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. +380 57 70-21-326, E-mail: Volodymyr.obrizan@gmail.com.

**Мищенко Александр Сергеевич**, аспирант кафедры АПВТ ХНУРЭ. Научные интересы: облачные технологии, web-программирование. Увлечения: путешествия. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. +380 57 70-21-326, E-mail: santific@gmail.com.

**Tamer Bani Amer**, магистрант ХНУРЭ. Научные интересы: проектирование и тестирование вычислительных систем. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. +380 57 70-21-326. E-mail: hahanov@kture.kharkov.ua.