

## Объективизация баз знаний интеллектуальных систем на основе индуктивного вывода с использованием нестрогих вероятностей

Аршинский Леонид Вадимович, Лебедев Вадим Сергеевич

Иркутский государственный университет путей сообщения,

Россия, Иркутск, *larsh@mail.ru*

**Аннотация.** Одним из способов объективизации продукционных баз знаний в системах, основанных на знаниях, может служить индуктивный вывод на основе объединённого метода сходства и различия с применением таблиц совместной встречаемости явлений. В статье предложен подход к такому выводу в условиях возможной низкой достоверности и противоречивости источников информации, формирующих таблицы. В основе подхода лежит понятие нестрогой вероятности, которая, в свою очередь, опирается на теорию логик с векторной семантикой в варианте  $V^{TF}$ -логик. Сами таблицы возможно получать из больших данных, в первую очередь – реляционных баз данных. Обсуждаемый подход позволяет не только генерировать продукции в условиях низкого качества информации, но и рассчитывать их истинность как векторную величину.

**Ключевые слова:** база знаний, большие данные, индуктивный вывод, нестрогая вероятность, логики с векторной семантикой

**Цитирование:** Аршинский Л.В. Объективизация баз знаний интеллектуальных систем на основе индуктивного вывода с использованием нестрогих вероятностей / Л.В. Аршинский, В.С. Лебедев // Информационные и математические технологии в науке и управлении. – 2022. – № 4(28). – С. 190-200. – DOI:10.38028/ESI.2022.28.4.015.

**Введение.** Системы, основанные на знаниях (СОЗ) – одна из классических технологий искусственного интеллекта (ИИ), с которой он фактически начинался. В основу технологии положено моделирование предметных областей (ПрО) с помощью баз знаний (БЗ), дополненных машинами вывода – специальными компонентами, обрабатывающими знания. Эти принципы были заложены ещё на заре развития ИИ в таких известных программах, как «Логик-теоретик» (Logic Theorist) Ньюэлла и Саймона, а также «Общий решатель проблем» (General Problem Solver - GPS) Ньюэлла, Саймона, Шоу [1, 2].

Моделирование на основе знаний позволяет формализовать области, не поддающиеся классической математической (количественной) формализации. Знаниевые модели качественные и отражают самые общие связи и отношения внутри ПрО. Уступая количественным моделям в детальности и предсказательной силе, они обладают большей шириной охвата, так как способны описывать любую часть мира, о которой можно рассуждать.

Первые СОЗ синтезировали или доказывали теоремы, исходя из соответствующих аксиом математики. Однако достаточно быстро стало понятно, что в таком качестве могут выступать определённые утверждения из любой предметной области: медицины, геологии, химии, юриспруденции и т.д. Здесь тоже существуют начальные факты и связи, отталкиваясь от которых, можно приходиться к тем или иным заключениям. Соответствующее направление ИИ получило название экспертных систем (ЭС) [3-5].

Источниками знаний – исходных утверждений о свойствах и отношениях ПрО – стали эксперты. Однако знания экспертов вторичны по отношению к ПрО, они отражают её, внося субъективизм в соответствующие модели. При том, что сами БЗ – это отражение знаний экспертов, где также присутствует субъективное знание, уже со стороны инженеров знаний (хотя и существуют специальные методы уменьшения этого). То есть, в отличие от классического математического моделирования, в СОЗ присутствует двойное отражение: ПрО – эксперт – БЗ. Это, безусловно, сказывается на качестве моделей. Кроме того, при отражении знаний экспертов вскрылся факт: «Чем более компетентными становятся эксперты, тем менее способны они описать знания, которые используют для решения задач» [4] (парадокс

экспертизы). Да и сами принципы человеческого мышления сильно отличаются от машинных рассуждений. Возникают и дополнительные риски, обусловленные влиянием человеческого фактора [6]. Всё это вызвало снижение интереса к ЭС и даже привело к крайним представлениям о неэффективности СОЗ, хотя речь идёт об естественных границах технологии.

Как показывает история ИИ, специалисты приложили немало усилий для расширения этих границ. Разрабатывались новые модели знаний, методы автоматизации рассуждений, включающие разнообразные логические и квазилогические формализмы; наряду со знаниями стали использовать метазнания, ЭС развились до гибридных ЭС, объединяющих стандартные пакеты прикладных программ и средства манипулирования знаниями [7]. Более того, появились гибридные интеллектуальные системы как технология, объединяющая различные методы ИИ, включая технологии СОЗ [8, 9]. Однако главным источником знаний для СОЗ по-прежнему остались эксперты, а значит, БЗ как модель ПрО неизбежно содержит изъян двойного отражения: ПрО – эксперт – БЗ, т.е. изъян субъективизма. Для уменьшения этого изъяна специалисты в качестве источников знаний предлагают привлекать не только экспертов, но и источники в виде лабораторных журналов, отчётов, иные первичные материалы, содержащие сведения непосредственно о ПрО [4, 10], в том числе базы данных (БД) [11]. Поскольку БД отражают непосредственно ПрО, становится возможным извлекать из них внутрипредметные связи и отношения, не искажённые субъективизмом экспертов-исследователей, причем в автоматическом или автоматизированном режимах. В информационных технологиях это обеспечивается средствами анализа больших данных (АБД) и интеллектуального анализа данных (Data Mining) как его ветви [10-14].

### 1. Индуктивный вывод в АБД. Векторная истинность при индуктивном выводе.

Сегодня АБД – одно из ведущих направлений в обработке информации. Накопленные на электронных носителях сведения становятся источником знаний, и «добыча» их из информационных массивов представляет большой прикладной интерес. Собственно, АБД во многом можно рассматривать, как способ извлечения скрытых знаний из больших массивов данных [11, 13].

Существуют различные подходы к анализу – статистические методы, OLAP-анализ, методы распознавания образов (задачи сегментации и классификации) и т.д. [13, 14] Естественное место в этом ряду занимает индуктивная логика, которая изначально предназначалась как раз для вывода общих закономерностей из частных случаев. В первую очередь – это закономерности вида «Если..., то...» [15-17]. Несмотря на то, что к индуктивному выводу в известном смысле можно отнести любой метод АБД, рассмотрим индукцию в классическом смысле Бэкона-Милля.

Первые упоминания об индукции, как методе, упоминаются ещё у Аристотеля и ряда более поздних исследователей [16]. Однако считается, что классический индуктивный вывод берёт своё начало в работах Ф. Бэкона и Д.С. Милля. Кроме самого первого – перечислительного метода, известного ещё Аристотелю, он содержит шесть «фигур» [16, 17]:

#### 1. Метод единственного сходства (МЕС):

$$\begin{array}{l} a_{11} \& \dots \& a_{n1} \& x \& y \\ a_{12} \& \dots \& a_{n2} \& x \& y \\ \dots \\ a_{1k} \& \dots \& a_{nk} \& x \& y \\ \hline x \rightarrow y \end{array}$$

#### 2. Метод единственного различия (МЕР):

$$\frac{a \& x \& y}{a \& \neg x \& \neg y}$$


---


$$x \rightarrow y$$

3. Объединённый метод сходства и различия, совмещающий МЕС и МЕР (ОМСР).  
 4. Метод сопутствующих изменений:

$$\frac{a \& x \& y}{a \& x' \& y'}$$


---


$$x \rightarrow y$$

5. Метод остатков:

$$\frac{a \& x \rightarrow b \& y}{a \rightarrow b}$$


---


$$x \rightarrow y$$

6. Обратная дедукция:

$$\frac{a \rightarrow b}{b}$$


---


$$a$$

В каждом случае заключение носит гипотетический характер. При этом фигуры 1-5 отражают свойство совместной встречаемости, или совместного изменения  $x$  и  $y$ ; что здесь является причиной, а что следствием, устанавливается из дополнительных соображений. С позиции дальнейшего интересны первые три фигуры, которые естественным образом ложатся на реляционную модель данных, наиболее часто употребляемую сегодня в различных БД. Эти фигуры фактически сводятся к анализу таблиц совместной встречаемости, подобных табл. 1.

**Таблица 1.** Совместная встречаемость явлений  $a_i$  и  $b$

	$a_1$	$a_2$	...	$a_n$	$b$
1	$a_{11}$	$a_{21}$	...	$a_{n1}$	$b_1$
2	$a_{12}$	$a_{22}$	...	$a_{n2}$	$b_2$
3	$a_{13}$	$a_{23}$	...	$a_{n3}$	$b_3$
...	...	...	...	...	...
$K$	$a_{1K}$	$a_{2K}$	...	$a_{nK}$	$b_K$

Здесь  $a_i$  и  $b$  – наблюдаемые явления, между которыми ищется связь типа  $a_i \rightarrow b$  (вид импликации может быть более сложным, но ограничимся таким). Факты наличия/отсутствия явлений рассматриваются двузначным (булевым) образом, когда  $a_{ik}, b_k \in \{0,1\}$  ( $k = 1 \dots K$ ,  $K$  – число опытов).

Анализ таблиц, подобных 1, даёт хороший результат, когда  $a_i$  и  $b$  совместно встречаются либо не встречаются во всех  $K$  позициях. Если не так (что бывает на практике) учитывают относительную долю случаев, когда связь подтверждается, потому данный подход дополнился статистико-вероятностными представлениями [15, 17-20]. Это позволило категорические суждения  $a_i \rightarrow b$  дополнить количественными характеристиками вроде показателей уверенности, как доле совместного наличия/отсутствия пар  $\{a_i, b\}$  в общем наборе опытов. Одна из существующих при этом проблем, на которую не всегда обращают внимание, – необходимость гарантированного присутствия или отсутствия элементов пары. Возможна ситуация, когда полная уверенность в результатах наблюдения отсутствует. Например, когда не выполнена или некачественно выполнена очистка данных, данные получены из разнородных, в т.ч. противоречащих друг другу источников, источникам нет полного доверия и т.д.

Популярный сегодня нечёткий подход здесь плохо применим, так как для него нет разницы между совместными ответами «Да и Нет» (противоречие) и «Не знаю» (неопределённость), притом в условиях неполного доверия к данным. Современные техники индуктивного вывода пробуют это учесть (например [20, 21]), но более естественным, как представляется, выглядит векторный подход, когда истинность суждения  $a$  (в нашем случае утверждения о наблюдении/ненаблюдении элементов пары  $\{a_i, b\}$ ) представляется вектором  $\|a\| = \langle a^+; a^- \rangle$ , где  $a^+, a^- \in [0,1]$ ;  $a^+$  – степень уверенности, что  $a$  истинно,  $a^-$ , – что ложно (аспекты истинности Истина и Ложь) [22, 23]. Аспекты  $a^+$  и  $a^-$  формируются подтверждающими и опровергающими источниками, которым мы можем в разной степени доверять или не доверять. При этом связь

$$a^+ + a^- = 1,$$

не постулируется. Как результат, таблица 1 принимает векторную форму (таблица 2).

**Таблица 2.** Векторное представление совместной встречаемости

	$a_1$	$a_2$	...	$a_n$	$b$
1	$\langle a_{11}^+; a_{11}^- \rangle$	$\langle a_{21}^+; a_{21}^- \rangle$	...	$\langle a_{n1}^+; a_{n1}^- \rangle$	$\langle b_1^+; b_1^- \rangle$
2	$\langle a_{12}^+; a_{12}^- \rangle$	$\langle a_{22}^+; a_{22}^- \rangle$	...	$\langle a_{n2}^+; a_{n2}^- \rangle$	$\langle b_2^+; b_2^- \rangle$
3	$\langle a_{13}^+; a_{13}^- \rangle$	$\langle a_{23}^+; a_{23}^- \rangle$	...	$\langle a_{n3}^+; a_{n3}^- \rangle$	$\langle b_3^+; b_3^- \rangle$
...	...	...	...	...	...
$K$	$\langle a_{1K}^+; a_{1K}^- \rangle$	$\langle a_{2K}^+; a_{2K}^- \rangle$	...	$\langle a_{nK}^+; a_{nK}^- \rangle$	$\langle b_K^+; b_K^- \rangle$

Здесь  $a_{ik}^+, b_k^+, a_{ik}^-, b_k^- \in [0,1]$ , так как по каждому из явлений  $a_i$  и  $b$  могут поступать как подтверждающие, так и опровергающие свидетельства. Переход к классическому случаю происходит, если векторы имеют значение  $\langle 1;0 \rangle$  (строгая истина) или  $\langle 0;1 \rangle$  (строгая ложь). Таблица 2 даёт основания к применению в индуктивном выводе понятия нестрогой вероятности.

**2. Нестрогая вероятность.** Данное понятие было введено в работе [23] и связано с векторным представлением истинности утверждения  $F(\omega, A) = \langle \text{«Элементарное событие } \omega \text{ благоприятно с точки зрения события } A \text{»} \rangle$ . Здесь  $A$  – подмножество полной группы элементарных событий  $\Omega$ . Если, как это принято, представить вероятность события  $A$  суммой:

$$p(A) = \sum_{\omega \in A} p(\omega), \tag{1}$$

то такой же результат даст сумма:

$$p(A) = \sum_{\omega \in \Omega} \|F(\omega, A)\| p(\omega), \tag{2}$$

где  $\|F(\omega, A)\|$  – истинность  $F(\omega, A)$ . Истинность равна 1, если  $\omega \in A$ , и 0 в противном случае. Переход к векторному представлению:

$$\|F(\omega, A)\| = \langle F^+(\omega, A); F^-(\omega, A) \rangle$$

порождает иное, векторное представление вероятности:

$$P(A) = \langle P^+(A); P^-(A) \rangle = \langle \sum_{\omega \in \Omega} F^+(\omega, A) p(\omega); \sum_{\omega \in \Omega} F^-(\omega, A) p(\omega) \rangle. \tag{3}$$

Оно имеет смысл, когда нет твердой уверенности в благоприятности/неблагоприятности  $\omega$  для  $A$ , а есть доводы как «за», так и «против» с разной степенью доверия к ним. Очевидно, что для строгих значений вектора  $\|F(\omega, A)\|$ , равных  $\langle 1;0 \rangle$  или  $\langle 0;1 \rangle$ , (2) превращается в (1), а (3) в привычную вероятность, где  $P^+(A)$  – вероятность  $A$ , а  $P^-(A)$  – вероятность противоположного события. В [10] для такого представления даны выражения для сложных вероятностей:

$$P(\neg A) = \sum_{\omega \in \Omega} \|\neg F(\omega, A)\| p(\omega) = \langle \sum_{\omega \in \Omega} F^-(\omega, A) p(\omega); \sum_{\omega \in \Omega} F^+(\omega, A) p(\omega) \rangle$$

– вероятность первой формы противоположного события;

$$P(\sim A) = \sum_{\omega \in \Omega} \|\sim F(\omega, A)\| p(\omega) = \langle \sum_{\omega \in \Omega} [1 - F^+(\omega, A)] p(\omega); \sum_{\omega \in \Omega} [1 - F^-(\omega, A)] p(\omega) \rangle$$

– вероятность второй формы противоположного события (в классическом случае эти формы совпадают);

$$P(A \vee B) = \langle \sum_{\omega \in \Omega} [F^+(\omega, A) \oplus F^+(\omega, B)] p(\omega); \sum_{\omega \in \Omega} [F^-(\omega, A) \bullet F^-(\omega, B)] p(\omega) \rangle$$

– вероятность первой формы суммы двух нестрогих событий;

$$P(A \vee_2 B) = \langle \sum_{\omega \in \Omega} [F^+(\omega, A) \oplus F^+(\omega, B)] p(\omega); \sum_{\omega \in \Omega} [F^-(\omega, A) \oplus F^-(\omega, B)] p(\omega) \rangle$$

– вероятность второй формы суммы двух нестрогих событий (в классическом случае эта вероятность отсутствует);

$$P(A \& B) = \langle \sum_{\omega \in \Omega} [F^+(\omega, A) \bullet F^+(\omega, B)] p(\omega); \sum_{\omega \in \Omega} [F^-(\omega, A) \oplus F^-(\omega, B)] p(\omega) \rangle$$

– вероятность первой формы произведения двух нестрогих событий;

$$P(A \&_2 B) = \langle \sum_{\omega \in \Omega} [F^+(\omega, A) \bullet F^+(\omega, B)] p(\omega); \sum_{\omega \in \Omega} [F^-(\omega, A) \bullet F^-(\omega, B)] p(\omega) \rangle$$

– вероятность второй формы произведения двух нестрогих событий (в классическом случае эта вероятность отсутствует).

Здесь  $\bullet$  и  $\oplus$  – соответственно триангулированная (треугольная) норма и ко-норма в инфиксной записи, с дополнительной аксиомой:

$$(1 - x) \bullet (1 - y) = 1 - x \oplus y;$$

(или, что то же самое:  $(1 - x) \oplus (1 - y) = 1 - x \bullet y$ ).

Двумя наиболее распространёнными примерами данных норм выступают известные пары функций:

$$\begin{aligned} x \bullet y &= x \cdot y; & x \oplus y &= x + y - x \cdot y; \\ x \bullet y &= \min(x, y); & x \oplus y &= \max(x, y). \end{aligned}$$

Смысл первых форм сложных событий достаточно очевиден. Для противоположного события благоприятность и неблагоприятность меняются местами. Для суммы достаточно, чтобы элементарное событие  $\omega$  было благоприятным хотя бы для одного из подмножеств  $A$  или  $B$  и неблагоприятным для обоих сразу. Для произведения – чтобы событие было благоприятным для обоих и неблагоприятным хотя бы для одного.

Вторые приведены для полноты и далее здесь не рассматриваются.

В [23] нестрогая вероятность интерпретировалась, как риск, если соответствующие случайные события оценивать, как с вредной, так и с полезной стороны. Возможны и другие интерпретации [24]. Одну из таких интерпретаций предлагает индуктивный вывод.

**3. Нестрогая вероятность в индуктивном выводе.** Как уже говорилось, в основе ОМСР лежит работа с таблицами совместной встречаемости (таблица 1). При этом каждая строка таблицы рассматривается, как аргумент в пользу соответствующей импликации (например,  $a_i \rightarrow b$ ). Согласно ОМСР, если каждое обнаружение явления  $a_i$  сопровождается регистрацией  $b$ , а отсутствие  $a_i$  отсутствием  $b$ , можно говорить о достоверности импликации (с точностью до произведённых опытов). Формально это означает эквивалентность  $a_i$  и  $b$ ; выполняется:  $a_i \& b \vee \neg a_i \& \neg b$  (в этом смысле «индуктивная импликация» отличается от материальной:  $\neg a_i \vee b$ ).

Если совместное наличие/отсутствие  $a_i$  и  $b$  наблюдается только в части проверок, скажем, в 75% случаев, можно говорить о степени уверенности в импликации равной 0.75. Каж-

дая строка таблицы 1, где одно из утверждений:  $a_i \& b$  или  $\neg a_i \& \neg b$  истинно, есть аргумент в пользу  $a_i \rightarrow b$ ; остальные строки – аргументы против.

Указанные представления можно перенести на векторный случай, когда утверждения о наблюдении не столь категоричны и в каждой строке могут встречаться как показания за, так и показания против  $a_i$  и/или  $b$  с разной степенью убедительности. Значение вектор-функции  $\|F(\omega, A)\|$  для конкретной строки  $k$  в этом случае определяется, как:

$$\|F(k, a_i \rightarrow b)\| = \|a_{ik} \& b_k \vee \neg a_{ik} \& \neg b_k\|;$$

(строка, как элементарное случайное событие).

Присваивая строке «вероятность»  $p(k)=1/K$  и принимая, что векторная истинность конъюнкции, дизъюнкции и отрицания определяются их первыми формами [23]:

$$\|A \& B\| = \langle A^+ \bullet B^+; A^- \oplus B^- \rangle;$$

$$\|A \vee B\| = \langle A^+ \oplus B^+; A^- \bullet B^- \rangle;$$

$$\|\neg A\| = \langle A^-; A^+ \rangle;$$

определяем для таблицы 2 истинность импликации  $a_i \rightarrow b$  как нестрогую вероятность:

$$P(a_i \rightarrow b) = \langle \frac{1}{K} \sum_{k=1}^K [a_{ik}^+ \bullet b_k^+ \oplus a_{ik}^- \bullet b_k^-]; \frac{1}{K} \sum_{k=1}^K [(a_{ik}^- \oplus b_k^-) \bullet (a_{ik}^+ \oplus b_k^+)] \rangle. \quad (4)$$

Выражение (4) можно использовать для вычисления (векторного) показателя уверенности (истинности) индукции  $a_i \rightarrow b$  при дефиците и/или противоречивости сведений о наблюдении  $a_i$  и  $b$ .

К примеру, для таблицы 3 (столбцы 2...n во внимание не принимаем; норму и ко-норму берём из первого примера):

**Таблица 3.** Пример совместной встречаемости 1

	$a_1$	$a_2$	...	$a_n$	$b$
1	$\langle 1;0 \rangle$	...	...	...	$\langle 0.8;0.4 \rangle$
2	$\langle 0.2;0.9 \rangle$	...	...	...	$\langle 0;1 \rangle$
3	$\langle 0.7;0.4 \rangle$	...	...	...	$\langle 1;0 \rangle$
4	$\langle 0.3;0.4 \rangle$	...	...	...	$\langle 1;1 \rangle$

Это даёт значение  $P(a_1 \rightarrow b) = \langle 0.745;0.5 \rangle$ . Т. е. импликация в целом наблюдается, но информация достаточно противоречивая. Если же взять более простые случаи (таблицы 4-6), получаются вполне ожидаемые результаты:

**Таблица 4.** Пример совместной встречаемости 2

	$a_1$	$a_2$	...	$a_n$	$b$
1	$\langle 1;0 \rangle$	...	...	...	$\langle 1;0 \rangle$
2	$\langle 0;1 \rangle$	...	...	...	$\langle 0;1 \rangle$
3	$\langle 1;0 \rangle$	...	...	...	$\langle 1;0 \rangle$
4	$\langle 0;1 \rangle$	...	...	...	$\langle 0;1 \rangle$

$P(a_1 \rightarrow b) = \langle 1;0 \rangle$  – строгая истина;

**Таблица 5.** Пример совместной встречаемости 3

	$a_1$	$a_2$	...	$a_n$	$b$
1	$\langle 1;0 \rangle$	...	...	...	$\langle 0;1 \rangle$
2	$\langle 0;1 \rangle$	...	...	...	$\langle 1;0 \rangle$
3	$\langle 0;1 \rangle$	...	...	...	$\langle 1;0 \rangle$
4	$\langle 0;1 \rangle$	...	...	...	$\langle 1;0 \rangle$

$P(a_1 \rightarrow b) = \langle 0.;1 \rangle$  – строгая ложь.

Если гипотеза подтверждается лишь в 75% случаев (таблица 6), получаем:

**Таблица 6.** Пример совместной встречаемости 4

	$a_1$	$a_2$	...	$a_n$	$b$
1	$\langle 1;0 \rangle$	...	...	...	$\langle 1;0 \rangle$
2	$\langle 0;1 \rangle$	...	...	...	$\langle 0;1 \rangle$
3	$\langle 1;0 \rangle$	...	...	...	$\langle 1;0 \rangle$
4	$\langle 1;0 \rangle$	...	...	...	$\langle 0;1 \rangle$

$P(a_1 \rightarrow b) = \langle 0.75;0.25 \rangle$  – как, собственно говоря, и должно быть.

Если считать, что подобные таблицы отражают связь между явлениями  $a$  и  $b$ , такая связь может выражаться гипотезами:

$$h_1 = a \rightarrow b = a \& b \vee \neg a \& \neg b$$

$$h_2 = a \rightarrow \neg b = a \& \neg b \vee \neg a \& b$$

$$h_3 = \neg a \rightarrow b = \neg a \& b \vee a \& \neg b$$

$$h_4 = \neg a \rightarrow \neg b = \neg a \& \neg b \vee a \& b$$

Видно, что  $h_1 = h_4$  и  $h_2 = h_3$ , то есть с точки зрения ОМСР возможны только гипотезы  $h_1$  и  $h_2$ .

Все сочетания «предельных» значений истинности  $\|a\|$  и  $\|b\|$ :  $\langle 0;1 \rangle$  «строгая ложь»,  $\langle 0;0 \rangle$  «неопределённость»,  $\langle 1;1 \rangle$  «полное противоречие»,  $\langle 1;0 \rangle$  «строгая истина» и соответствующие значения истинности гипотез  $h_1$  и  $h_2$  представлены в таблице 7. Видно, что для «классического» представления истинности в виде строгой истины и строгой лжи гипотезы  $h_1$  и  $h_2$  противоположны (таблица 7), то есть в «классическом» варианте достаточно одной гипотезы:  $h_1$  или  $h_2$ . Однако если  $\|a\|$  и  $\|b\|$  отличаются от предельных значений, такая взаимосвязь исчезает:

$$h_1 = \langle a^+ \bullet b^+ \oplus a^- \bullet b^-; (a^- \oplus b^-) \bullet (a^+ \oplus b^+) \rangle$$

$$h_2 = \langle a^+ \bullet b^- \oplus a^- \bullet b^+; (a^- \oplus b^+) \bullet (a^+ \oplus b^-) \rangle$$

Это означает, что в векторном случае свидетельства в пользу каждой гипотезы следует учитывать отдельно!

**Таблица 7.** Гипотезы  $h_1$  и  $h_2$  для предельных значений истинности

$a$	$b$	$a \rightarrow b$	$a \rightarrow \neg b$
$\langle 0;1 \rangle$	$\langle 0;1 \rangle$	$\langle 1;0 \rangle$	$\langle 0;1 \rangle$
$\langle 0;0 \rangle$	$\langle 0;1 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$
$\langle 1;1 \rangle$	$\langle 0;1 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$
$\langle 1;0 \rangle$	$\langle 0;1 \rangle$	$\langle 0;1 \rangle$	$\langle 1;0 \rangle$
$\langle 0;1 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$
$\langle 0;0 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$
$\langle 1;1 \rangle$	$\langle 0;0 \rangle$	$\langle 0;1 \rangle$	$\langle 0;1 \rangle$
$\langle 1;0 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$
$\langle 0;1 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$
$\langle 0;0 \rangle$	$\langle 1;1 \rangle$	$\langle 0;1 \rangle$	$\langle 0;1 \rangle$

$a$	$b$	$a \rightarrow b$	$a \rightarrow \neg b$
$\langle 1;1 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$
$\langle 1;0 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$
$\langle 0;1 \rangle$	$\langle 1;0 \rangle$	$\langle 0;1 \rangle$	$\langle 1;0 \rangle$
$\langle 0;0 \rangle$	$\langle 1;0 \rangle$	$\langle 0;0 \rangle$	$\langle 0;0 \rangle$
$\langle 1;1 \rangle$	$\langle 1;0 \rangle$	$\langle 1;1 \rangle$	$\langle 1;1 \rangle$
$\langle 1;0 \rangle$	$\langle 1;0 \rangle$	$\langle 1;0 \rangle$	$\langle 0;1 \rangle$

Из таблицы 7 также видно, что, несмотря на то, что результирующие истинности достаточно разумны, строки с неопределённостями целесообразно исключать. Исключение можно реализовать, вводя порог определённости. Например, 0.5, если его значение  $\in [0,1]$ . Окончательный выбор следует делать, привлекая меры достоверности:

$$\mu_d(h_k) = h_k^+ - h_k^-;$$

и определённости:

$$\mu_o(h_k) = h_k^+ \oplus h_k^-.$$

Первая из них позволяет выбрать наиболее достоверную гипотезу, вторая – определить гипотезу, наиболее подкреплённую свидетельствами.

Таким образом, подход на основе нестрогих вероятностей позволяет определять (статистическую) истинность индуктивных гипотез как векторную величину, независимо от наличия или отсутствия проблем со свидетельствами. Достоинством подхода служит то, что техника расчёта сохраняет работоспособность при наличии описанных проблем и в этом смысле устойчива к ним. Объективные имплицативные связи можно получать в достаточно сложных условиях. Однако субъективизм может (и скорее всего будет) присутствовать в векторных оценках истинности фактов (наборы  $a_{ik}^+, b_k^+, a_{ik}^-, b_k^-$ ), что следует принимать во внимание.

**Заключение.** Общий вывод следующий:

1. Индуктивный вывод в смысле ОМСР в условиях дефицита и противоречивости данных можно осуществлять на основе векторного представления истинности и вытекающего из него понятия нестрогой вероятности.

2. Такой подход позволяет работать в условиях частичной информированности и, что важно, автоматически и достаточно последовательно учитывать неполноту и противоречивость свидетельств, отсеивая их при необходимости. Это выглядит достоинством представленного подхода, хотя следует помнить, что субъективизм может присутствовать в оценках истинности генерируемых импликаций.

3. Саму технику можно применять для автоматизации построения продукционных баз знаний в СОЗ. Исходным сырьём для этого могут служить большие данные.

Следует отметить, что ОМСР не единственный способ выявления причинно-следственных зависимостей в наборах данных. Этому же служат, к примеру, ассоциативный и секвенциальный анализы, направленные на выявление связей между неупорядоченными (в ассоциативном) или упорядоченными (в секвенциальном анализе) группами явлений или объектов, входящих в т.н. транзакции по принципу: если в транзакции наблюдается ассоциация/секвенция  $A$ , то с определённой степенью уверенности в ней может обнаружиться ассоциация/секвенция  $B$  (см. напр. [25, 26]). Основным понятием обоих анализов является «поддержка» как отношение числа транзакций, где ассоциация/секвенция (или совместное присутствие  $A$  и  $B$ ) наблюдалась, к общему числу транзакций. Например, вывод о том, что если

покупатель приобрёл группу товаров *A*, в его чеке (транзакции) будет наблюдаться и группа товаров *B*. Если полный перечень (ассортимент) явлений представить набором столбцов, а транзакции строками с нулями и единицами в соответствующих столбцах, то поддержку можно вычислить (и вычисляют) статистически. Заменяя нули и единицы векторами истинности, можно распространить представленный в работе подход и на эти анализы. Это может оказаться востребованным для случаев, когда сведения о явлениях в транзакциях неочевидны, однако это требует отдельного изучения.

### Список источников

1. Newell A., Simon H.A. The Logic Theory Machine. Transactions on Information Theory, IT-2, no. 3, 1956.
2. Newell A., Shaw J.C., Simon H.A. Report on a General Problem-Solving Program. Carnegie Institute of Technology, 1958, 27 p.
3. Элти Дж. Экспертные системы: концепции и примеры / Дж. Элти, М. Кумбс: пер. с англ. и предисл. Б.И. Шитикова // М.: Финансы и статистика, 1987. – 191 с.
4. Уотермен Д. Руководство по экспертным системам / Д. Уотермен: пер. с англ. // М.: Мир, 1989. – 388 с.
5. Джарратано Дж. Экспертные системы: принципы разработки и программирование. 4-е издание. / Дж. Джарратано, Г. Райли: пер. с англ. // М.: Издательский дом «Вильямс», 2007. – 1152 с.
6. Абрамова Н.А. О проблеме рисков из-за человеческого фактора в экспертных методах и информационных технологиях / Н.А. Абрамова // Труды VI Международной конференции «Идентификация систем и задачи управления» SICPRO`07 Москва 29 января-1 февраля 2007 г. – С. 51-68.
7. Частиков А.П. Разработка экспертных систем. Среда CLIPS / А.П. Частиков А.П., Т.А. Гаврилова Т.А., Д.Л. Белов // СПб: «БХВ-Петербург», 2003. – 393 с.
8. Medsker L.R. Hybrid Intelligent Systems. Kluwer Academic Publishers, Boston, London, Dordrecht, 1995, 298 p.
9. Колесников А.В. Технология гибридных интеллектуальных систем: специальность 05.13.01 «Системный анализ, управление и обработка информации»: диссертация на соискание учёной степени доктора технических наук / А.В. Колесников // Санкт-Петербургский государственный технический университет. – Санкт-Петербург, 2002. – 387 с.
10. Гаврилова Т.А. Инженерия знаний. Модели и методы: учебник / Т.А. Гаврилова, Д.В. Кудрявцев, Д.И. Муромцев // СПб.: Издательство «Лань», 2016. – 324 с.
11. Финн В.К. Об интеллектуальном анализе данных / В.К. Финн // Новости искусственного интеллекта, 2004. – №3. – С. 1-20.
12. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко // Новосибирск: ИМ СО РАН, 1999. – 270 с.
13. Большие данные (Big Data) // ФОРС. Интернет-журнал, № 1. URL: [https://www.fors.ru/upload/magazine/01/html\\_texts/total\\_big\\_date\(2\).html](https://www.fors.ru/upload/magazine/01/html_texts/total_big_date(2).html)
14. Big Data Analytics – Quick Guide. Available at: [https://www.tutorialspoint.com/big\\_data\\_analytics/big\\_data\\_analytics\\_quick\\_guide.htm](https://www.tutorialspoint.com/big_data_analytics/big_data_analytics_quick_guide.htm)
15. Пойа Д. Математика и правдоподобные рассуждения / Д. Пойа: пер. с англ. под ред. С.А. Яновской // М.: Наука, 1975. – 464 с
16. Ивлев Ю.В. Логика: Учебник 3-е изд. / Ю.В. Ивлев // М.: ТК Велби, – Изд-во Проспект, 2004. – 288 с.
17. Минто В. Дедуктивная и индуктивная логика / В. Минто // Мн.: Харвест, 2002. – 352 с.
18. Голенков В.В. Статистические основы индуктивного вывода: учеб. пособие / В.В. Голенков, М. Д. Степанова, С.А. Самодумкин, Н.А. Гулякина // Минск: БГУИР, 2009. – 202 с.
19. Кайберг Г. Вероятность и индуктивная логика / Г. Кайберг // М.: Изд-во «Прогресс», 1978. – 373 с.
20. Гаек П., Гавранек Т. Автоматическое образование гипотез: математические основы общей теории / П. Гаек, Т. Гавранек // М.: Наука. Главная редакция физико-математической литературы, 1984. – 280 с.
21. ДСМ-метод автоматического порождения гипотез: Логически и эпистемологические основания / Сост. О.М. Аншаков, Е.Ф. Фабрикантова // М.: Книжный дом «ЛИБРИКОМ», 2009. – 432 с.
22. Аршинский Л.В. Методы обработки нестрогих высказываний / Л.В. Аршинский. // Иркутск: Изд-во Восточно-Сибирского института МВД России, 1998. – 40 с.
23. Аршинский Л.В. Приложение логик с векторной семантикой к описанию случайных событий и оценке риска / Л.В. Аршинский. // Проблемы анализа риска, 2005. –Т.2. – № 3. – С.231-248.
24. Аршинский Л.В. Векторные логики: основания, концепции, модели / Л.В. Аршинский // Иркутск: Иркут. гос. ун-т, 2007. – 228 с.

25. Трубицын И.Ю. Формирование наборов данных при проведении секвенциального анализа событий / И.Ю. Трубицын, Я.А. Бекенёва // Известия СПбГЭТУ «ЛЭТИ», 2020. – № 3 – С. 45-52.

26. Матвейкин В.Г., Дмитриевский Б.С., Ляпин Н.Р. Информационные системы интеллектуального анализа / В.Г. Матвейкин, Б.С. Дмитриевский, Н.Р. Ляпин // М.: Машиностроение, 2008. – 92 с.

*Аршинский Леонид Вадимович. Д.т.н., доцент, профессор кафедры «Информационные системы и защита информации» Иркутского государственного университета путей сообщения, AuthorID: 520252; SPIN: 9286-4084; ORCID: 0000-0001-5135-7921, larsh@mail.ru, Россия, Иркутск, Чернышевского, 15.*

*Лебедев Вадим Сергеевич. Аспирант кафедры «Информационные системы и защита информации» Иркутского государственного университета путей сообщения. AuthorID: 1100520; SPIN: 4276-7301. lebedevvs97@yandex.ru, Россия, Иркутск, Чернышевского, 15.*

UDC 004.89+510.644

DOI:10.38028/ESI.2022.28.4.015

## Objectification of knowledge bases of intelligent systems based on inductive inference using non-strict probabilities

Leonid V. Arshinskiy, Vadim S. Lebedev

Irkutsk State Transport University,

Russia, Irkutsk, *larsh@mail.ru*

**Abstract.** It is shown that one of the ways to objectify productive knowledge bases in knowledge-based systems can serve as an inductive inference based on the combined method of similarity and difference and tables of joint occurrence of phenomena. An approach to the use of such a conclusion is proposed in conditions of possible low reliability and inconsistency of information sources forming tables. The approach is based on the concept of non-strict probability, which, in turn, is based on the theory of logics with vector semantics in the VTF-logic variant. The tables themselves can be obtained from big data, primarily relational databases. It is assumed that such an approach will weaken subjectivism in the construction of production knowledge bases.

**Keywords:** knowledge base, big data, inductive inference, nonstrict probability, logics with vector semantics

### References

1. Newell A., Simon H.A. The logic theory machine. Transactions on information theory, IT-2, no. 3, 1956.
2. Newell A., Shaw J.C., Simon H.A. Report on a General problem-solving program. Carnegie Institute of Technology, 1958, 27 p.
3. Alty J.L., Coombs M.J. Ekspertnyye sistemy: kontseptsii i primery [Expert Systems: Concepts and Examples]. Moscow: Finansy i statistika [Finance and statistics], 1987, 191 p.
4. Uotermen D. Rukovodstvo po ekspertnym sistemam [Guide to Expert Systems]. Moscow: Mir [Mir], 1989, 388 p.
5. Giarratano J.C., Riley G.D. Ekspertnye sistemy: principy razrabotki i programmirovaniye. 4-e izdanie [Expert Systems: Principles and Programming. 4th ed.]. Moscow: Izdatel'skiy dom «Vil'yams» [Publishing House "Williams"], 2007, 1152 p.
6. Abramova N.A. O probleme riskov iz-za chelovecheskogo faktora v ekspertnyh metodah i informacionnyh tekhnologiyah [On the problem of risks due to the human factor in expert methods and information technologies]. Trudy VI Mezhdunarodnoy konferentsii «Identifikatsiya sistem i zadachi upravleniya» SICPRO`07 Moskva 29 yanvarya-1 fevralya 2007g. [Proceedings of the VI International Conference "Identification of systems and management tasks" SICPRO`07 Moscow January 29-February 1], 2007, pp. 51-68.
7. Chastikov A.P., Gavrilova T.A., Belov D.L. Razrabotka ekspertnyh sistem. Sreda CLIPS. [Development of expert systems. CLIPS environment]. St. Petersburg: "BHV-Peterburg" ["BHV-Petersburg"], 2003, 393 p.
8. Medsker L.R. Hybrid Intelligent Systems. Kluwer Academic Publishers, Boston, London, Dordrecht, 1995, 298 p.
9. Kolesnikov A.V. Tekhnologiya gibridnyh intellektual'nyh system [Technology of hybrid intelligent systems]: specialty 05.13.01 "System analysis, control and information processing": dissertation for the degree of Doctor of Technical Sciences. St. Petersburg State Technical University. Saint Petersburg, 2002, 387 p.
10. Gavrilova T.A., Kudryavtsev D.V., Muromtsev D.I. Inzheneriya znaniy. Modeli i metody: uchebnik [Engineering of knowledge. Models and methods: textbook]. St. Petersburg: Izdatel'stvo "Lan" [Publishing House "Lan"], 2016, 324 p.

11. Finn V.K. Ob intellektual'nom analize dannyh [About intelligent data analysis]. Novosti iskusstvennogo intellekta [Artificial Intelligence News], 2004, no. 3, pp. 1-20.
12. Zagoruiko N.G. Prikladnye metody analiza dannyh i znaniy [Applied methods of data and knowledge analysis]. Novosibirsk: IM SO RAN [MI SB RAS], 1999, 270 p.
13. Bol'shie dannye [Big Data]. FORS. Internet-zhurnal [FORS. Internet-magazine], no. 1. Available at: [https://www.fors.ru/upload/magazine/01/html\\_texts/total\\_big\\_date\(2\).html](https://www.fors.ru/upload/magazine/01/html_texts/total_big_date(2).html)
14. Big Data Analytics – Quick Guide. Available at: [https://www.tutorialspoint.com/big\\_data\\_analytics/big\\_data\\_analytics\\_quick\\_guide.htm](https://www.tutorialspoint.com/big_data_analytics/big_data_analytics_quick_guide.htm)
15. Poja G. Matematika i pravdopodobnye rassuzhdeniya [Mathematics and Plausible Reasoning]. Moscow: Nauka [Science], 1975, 464 p.
16. Ivlev Yu.V. Logika: Uchebnik. 3-e izd. [Logic: Textbook. 3rd ed.]. Moscow: TK Velbi [TK Velbi], Izd-vo Prospekt [Publishing House "Prospect"], 2004, 288 p.
17. Minto V. Deduktivnaya i induktivnaya logika [Deductive and inductive logic]. Minsk: Harvest [Harvest], 2002, 352 p.
18. Golenkov V.V., Stepanova M.D., Samodumkin S.A., Gulyakina N.A. Statisticheskie osnovy induktivnogo vyvoda: ucheb. posobie [Statistical bases of inductive inference: textbook]. Minsk: BGUIR [BGUIR], 2009, 202 p.
19. Kyburg H.E. Вероятность и индуктивная логика [Probability and Inductive Logic]. Moscow: Izd-vo «Progress» [Publishing House "Progress"], 1978, 373 p.
20. Hajek P., Havranek T. Avtomaticheskoye obrazovaniye gipotez: matematicheskiye osnovy obshchey teorii [Mechanizing Hypothesis Formation: Mathematical Foundation for a General Theory]. Moscow: Nauka. Glavnaya redakciya fiziko-matematicheskoy literatury [Science. Main editorial office of the physical and mathematical literature], 1984, 280 p.
21. Anshakov O.M., Fabrikantova E.F. DSM-metod avtomaticheskogo porozhdeniya gipotez: Logicheski i epistemologicheskie osnovaniya [DSM-the method of automatic generation of hypotheses: Logical and epistemological foundations]. Moscow: Knizhnyj dom "LIBRIKOM" ["LIBRIKOM" Book House], 2009, 432 p.
22. Arshinskiy L.V. Metody obrabotki nestrogih vyskazyvaniy [Methods of processing of nonstrict propositions]. – Irkutsk: Izd-vo Vostochno-Sibirskogo instituta MVD Rossii [East-Siberian Institute of MIA of Russia], 1998, 40 p.
23. Arshinskiy L.V. Prilozhenie logik s vektornoj semantikoj k opisaniyu sluchajnyh sobytij i ocenke riska [Application of logic with vector semantics to the description of random events and risk assessment]. Problemy analiza riska [Issues of Risk Analysis], 2005, vol.2, no. 3, pp. 231-248.
24. Arshinskiy L.V. Vektornye logiki: osnovaniya, koncepcii, modeli [Vector logic: foundations, concepts, models]. Irkutsk: Irkutskij gosudarstvennyj universitet [Irkutsk state university], 2007, 228 p.
25. Trubityn I. Yu., Bekeneva Ya, A. Formirovanie naborov danny`x pri provedenii sekvencial`nogo analiza soby`tij [Formation of databases during sequential event analysis]. Izvestiya SPbGETU «LETI» [Proceedings of Saint Petersburg Electrotechnical University], 2020, no. 3, pp. 45-52.
26. Matveikin V.G., Dmitrievsky B.S., Lyapin N.R. Informacionny`e sistemy` intellektual`nogo analiza [Information systems of intellectual analysis]. Moscow: Mashinostroenie [Mechanical engineering], 2008, 92 p.

**Leonid Vadimovich Arshinsky.** Doctor of Technical Sciences, Associate Professor, Professor of the Department "Information Systems and Information Security" of Irkutsk State Transport University, AuthorID: 520252, SPIN: 9286-4084, ORCID: 0000-0001-5135-7921, [larsh@mail.ru](mailto:larsh@mail.ru), Russia, Irkutsk, Chernyshevsky, 15.

**Vadim Sergejevich Lebedev.** Postgraduate student of the Department "Information Systems and Information Security" of Irkutsk State Transport University, AuthorID: 1100520, SPIN: 4276-7301, [lebedevs97@yandex.ru](mailto:lebedevs97@yandex.ru), Russia, Irkutsk, Chernyshevsky, 15.

Статья поступила в редакцию 03.09.2022; одобрена после рецензирования 22.09.2022; принята к публикации 01.11.2022.

The article was submitted 09/03/2022; approved after reviewing 09/22/2022; accepted for publication 11/01/2022.