Библиографические ссылки

1. Kennel M. B., Brown R., Abarbanel H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction // Phys. Rev. A. 1992. Vol. 45. № 6. P. 3403–3411.

2. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors // Physica D. 1983. Vol. 9. P. 189.

3. An unusually active sun during recent decades compared to the previous 11 000 years / S. Solanki, I. Usoskin, B. Kromer et al. // Nature. 2004. Vol. 431. № 7012. P. 1084–1087.

M. A. Bondarenko

METHODS FOR CHAOTIC DYNAMICS IN THE PROBLEM OF ANALYSIS OF WOLFER'S TIME SERIES

The results of computer studies of time series composed of the Wolfer's numbers are presented in the article. In accordance with the approach based on Takens's theorem, the author solves the problem of reconstructing dimension of attractor of a dynamical system describing the dynamics of the Sun. It is shown that for the last 3 000 years, fundamental changes in the evolution of the Sun have not happened.

Keywords: Taken's theorem, reconstruction of the attractor, Wolfer's numbers.

УДК 519.67-519.24, 519.6

А. А. Викентьев

О ВВЕДЕНИИ МЕТРИК НА ВЫСКАЗЫВАНИЯХ ЭКСПЕРТОВ С ВЕРОЯТНОСТЯМИ*

Предлагаются способы введения метрик на высказываниях экспертов, заданных формулами исчисления высказываний с вероятностями. Результаты переносятся на формулы над бесконечными носителями.

Ключевые слова: высказывания экспертов, кластеризация, расстояния, метрики.

В настоящее время проявляется все больший интерес к анализу экспертной информации, заданной в виде вероятностных логических высказываний экспертов. Интересны исследования о высказываниях экспертов, представленных формулами исчисления высказываний (ИВ) с вероятностями. Возникают задачи об алгоритмах распознавания закономерностей, согласования логических (экспертных) знаний и их кластеризации [1–6]. Для решения этих задач необходимы метрики на знаниях.

В данной статье рассматриваются логические высказывания экспертов, представленные формулами исчисления высказываний с вероятностями, предлагаются способы задания расстояний на формулах-высказываниях с вероятностями, а также устанавливаются свойства введенных расстояний, для чего используются вероятностный и теоретико-модельный подходы [2; 5; 7].

Основные определения. Будем рассматривать знания экспертов, представленные формулами ИВ с вероятностями (вероятностные высказывания), т. е. высказывания вида «ϕ с вероятностью p_ϕ», где ϕ – формула ИВ. Для таких высказываний будем использовать следующую запись:

$$B_i = \langle \phi, p_\phi \rangle, \quad B_j = \langle \psi, p_\psi \rangle.$$

Пусть Σ – база знаний, состоящая из формул ИВ, т. е. в Σ содержатся все формулы, с которыми будут работать эксперты; $S(\phi)$ – носитель формулы ϕ, т. е. множество элементарных высказываний, используемых при написании формулы ϕ, $S(\Sigma) = \bigcup_{\phi \in \Sigma} S(\phi)$ – носитель совокупности знаний.

Рассмотрим совокупность $P(S(\Sigma)) = 2^{S(\Sigma)}$ – множество всевозможных подмножеств множества $S(\Sigma)$. Для простоты обозначения примем, что $|P(S(\Sigma))| = 2^{|S(\Sigma)|} = n.$ Элементы множества $P(S(\Sigma))$ называются *моделями*. Более подробно о теории моделей см. в [2–4].

Пусть эксперты говорят о вероятностях (частости) формул на множестве всех n моделей, и каждое высказывание присутствует только с одной вероятностью. Тогда будем интерпретировать вероятность, данную экспертом, следующим образом: $B = \langle \phi, p_\phi \rangle$. Это означает, что высказывание φ истинно на $n_\phi = [\, n \cdot p_\phi \,]$ моделях, где $n = 2^{|S(\Sigma)|}$ – число всех моделей.

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты 07-01-00331а, 08-07-00136а).

Пусть имеется два вероятностных логических высказывания $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$. Дадим способ вычисления расстояния $\rho(B_i, B_j)$ между такими высказываниями.

Интерпретируя данные экспертами вероятности описанным выше способом, получаем, что высказывание ϕ истинно на $n_\phi = \left[n \cdot p_\phi \right]$ моделях, а высказывание ψ – на $n_\psi = \left[n \cdot p_\psi \right]$ моделях. Отметим, однако, что при таком подходе мы не знаем, на каких именно моделях каждое высказывание истинно, а также число моделей, на которых эти высказывания истинны одновременно.

Решим следующую задачу: пусть высказывание ϕ истинно на n_ϕ моделях, высказывание ψ истинно на n_ψ моделях и k – число моделей, на которых эти высказывания истинны одновременно. Вычислим расстояние между высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$.

Обозначим рассматриваемые далее расстояния через $\rho_k(B_i, B_j)$, где $k = 0, 1, 2, \ldots, \min(n_\phi, n_\psi)$.

Как и в [2–4], расстояние $\rho_k(B_i, B_j)$ для каждого $k = 0, 1, 2, \ldots, \min(n_\phi, n_\psi)$ определим через разность:

$$\rho_k(B_i, B_j) = \frac{n_\phi - k + n_\psi - k}{n} = \frac{n_\phi + n_\psi - 2k}{n}.$$

Теорема 1. Для расстояний $\rho_k(B_i, B_j)$ справедливы свойства:

1) $0 \le \rho_k(B_i, B_j) \le 1$;

2) $\rho_k(B_i, B_j) = \rho_k(B_j, B_i)$;

3) $\rho_k(B_i, B_j) \le \rho_k(B_i, B_s) + \rho_k(B_s, B_j)$;

4) $B_i \equiv B_j \Leftrightarrow \rho_k(B_i, B_j) = 0$ ($B_i \equiv B_j \Leftrightarrow \varphi \equiv \psi$ и $p_\phi = p_\psi$, т. е. формулы φ и ψ истинны на одних и тех же моделях);

5) $B_i \equiv {}^\neg B_j \Leftrightarrow \rho_k(B_i, B_j) = 1$;

6) $\rho_k(B_i, B_j) = 1 - \rho_k(B_i, {}^\neg B_j) = \rho_k({}^\neg B_i, {}^\neg B_j)$;

7) $\rho_k(B_i, B_j) = \rho_k(B_i \wedge B_j, B_i \vee B_j)$.

Докажем неочевидное свойство 3. Определение $\rho_k(B_i, B_j)$ можно переписать следующим образом:

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n} = \frac{n_{\phi\Delta\psi}}{n} = \frac{n_{(\neg\phi\wedge\psi)\vee(\phi\wedge\neg\psi)}}{n}$$

(по определению симметрической разности). Тогда для произвольного высказывания $B_s = \langle \chi, p_\chi \rangle$ нетрудно доказать, что $n_{\phi\Delta\psi} \le n_{\phi\Delta\chi} + n_{\chi\Delta\psi}$, откуда

$$\rho_k(B_i, B_j) = \frac{n_{\phi\Delta\psi}}{n} \le \frac{n_{\phi\Delta\chi}}{n} + \frac{n_{\chi\Delta\psi}}{n} = \\ = \rho_k(B_i, B_s) + \rho_k(B_s, B_j).$$

Докажем свойство 4 сначала слева направо ($\Rightarrow$). Если $B_i \equiv B_j$, то $\phi \equiv \psi$ и, значит, $n_\phi = n_\psi = k$. Следовательно,

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n} = 0.$$

Докажем свойство 4 в обратную сторону ($\Leftarrow$). Если $\rho_k(B_i, B_j) = 0$, то $n_\phi + n_\psi - 2k = 0$. Так как k может принимать значения $0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$, то $(n_\phi + n_\psi - 2k = 0 \Leftrightarrow n_\phi = n_\psi = k)$. Следовательно, $\phi \equiv \psi$ и, значит, $B_i \equiv B_j$.

Докажем свойство 5 слева направо ($\Rightarrow$). Если $B_i \equiv {}^\neg B_j$, то $\phi \equiv {}^\neg\psi$. Тогда $n_\phi = n - n_\psi$ и $k = 0$, следовательно

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 0}{n} = \frac{n}{n} = 1.$$

Докажем свойство 5 в обратную сторону ($\Leftarrow$). Если $\rho_k(B_i, B_j) = 1$, то $n_\phi + n_\psi - 2k = n$. Так как k может принимать значения $0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$, то $(n_\phi + n_\psi - 2k = n \Leftrightarrow n_\phi + n_\psi = n$ и $k = 0)$. Следовательно, $\phi \equiv {}^\neg\psi$ и $B_i \equiv {}^\neg B_j$.

Докажем свойство 6:

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n},$$

$$\rho_k(B_i, {}^\neg B_j) = \frac{n_\phi + n_{\neg\psi} - 2(n_\phi - k)}{n} = \frac{n_\phi + (n - n_\psi) - 2(n_\phi - k)}{n} = \\ = \frac{n - n_\phi - n_\psi - 2k}{n} = 1 - \frac{n_\phi + n_\psi - 2k}{n} = 1 - \rho_k(B_i, B_j),$$

Откуда нетрудно доказать, что

$$1 - \rho_k(B_i, {}^\neg B_j) = \rho_k({}^\neg B_i, {}^\neg B_j).$$

Докажем свойство 7.

Легко доказать, что $n_{(\phi\wedge\psi)\Delta(\phi\vee\psi)} = n_{\phi\Delta\psi}$. Тогда

$$\rho_k(B_i \wedge B_j, B_i \vee B_j) = \frac{n_{(\phi\wedge\psi)\Delta(\phi\vee\psi)}}{n} = \frac{n_{\phi\Delta\psi)}}{n} = \rho_k(B_i, B_j).$$

Далее предложим несколько способов вычисления расстояния $\rho(B_i, B_j)$ между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$.

Так как нам неизвестно число k, т. е. число моделей, на которых высказывания ϕ и ψ истинны одновременно, и если нет никаких предпочтений для значения k (хотя они и могут быть высказаны экспертами), то мы можем поступить следующим образом.

Предположим, что все значения числа k равновероятны. Тогда расстояние между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ определим как усреднение расстояний $\rho_k(B_i, B_j)$ по всем значениям k:

$$\rho(B_i, B_j) = \frac{\sum_{k=0}^{\min(n_\phi, n_\psi)} \rho_k(B_i, B_j)}{\min(n_\phi, n_\psi) + 1}.$$

Для этого расстояния также справедлива теорема 1, и слагаемые под знаком суммы можно взять с весами, в которых учтены мнения экспертов.

Если экспертами указано, какое значение k предпочтительнее, то в качестве $\rho(B_i, B_j)$ берется расстояние $\rho_k(B_i, B_j)$. Это возможно в случае, когда мы знаем, что пересечение состоит из k моделей.

Можно подойти к вопросу определения расстояния $\rho_k(B_i, B_j)$ и с вероятностно-статистической точки зрения и для каждого k вычислить частоту того, что высказывания ϕ и ψ одновременно истинны на k моделях.

Найдем частоту (вероятность) p_k того, что в выбранных n_ϕ и n_ψ моделях (они выбираются из n моделей) будет k моделей, на которых высказывания ϕ и ψ истинны одновременно, где $k = 0, 1, 2, \ldots, K,$ $\min(n_\phi, n_\psi)$.

Сначала определим вероятностное пространство $\langle \Omega, A, p \rangle$, где Ω – пространство элементарных исходов – моделей, $\Omega = \{0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)\}$ – число возможных совпадений моделей в наборах из n_ϕ и n_ψ моделей; A – система пар подмножеств множества моделей Ω, образующая σ-алгебру событий; p – вероятность на $\langle \Omega, A \rangle$.

Определим для $k = 0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$ на декартовом произведении $\Omega \otimes A$ случайную величину ξ:

$$\xi(k) = \rho_k(B_i, B_j),$$

т. е. функция ξ каждому k из Ω будет ставить в соответствие число – расстояние $\rho_k(B_i, B_j)$.

Вероятность этого события (с пересечением k) на классе из n моделей можно вычислить так:

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}}{C_n^{n_\phi} C_n^{n_\psi}},$$

где $C_n^{n_\phi}$ – число способов выбрать n_ϕ моделей из n моделей.

Действительно, любой набор, состоящий из n_ϕ моделей, может сочетаться с любым набором, состоящим из n_ψ моделей, значит число $(C_n^{n_\phi} \cdot C_n^{n_\psi})$ определяет количество всех способов выбрать два набора, один из которых состоит из n_ϕ моделей, а другой – из n_ψ моделей. Выбрать k моделей, которые будут общими в этих наборах, из n моделей можно C_n^k способами. Тогда остальные $(n_\phi - k)$ и $(n_\psi - k)$ моделей должны быть дизъюнктными. Следовательно, остальные $(n_\phi - k)$ моделей для пополнения набора, состоящего из k, до n_ϕ моделей можно выбрать $C_{n-k}^{n_\phi - k}$ способами, а $(n_\psi - k)$ моделей для получения набора, состоящего из n_ψ моделей, с учетом наших предположений, – $C_{n-n_\phi}^{n_\psi - k}$ способами. Итак, имеется всего $\left(C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}\right)$ способов выбрать два набора, один из которых состоит из n_ϕ моделей, а другой – из n_ψ моделей, и k моделей в этих наборах моделей совпадут. Поэтому вероятность того, что k моделей совпадет в наборах из n_ϕ и n_ψ элементных моделей, составит

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}}{C_n^{n_\phi} C_n^{n_\psi}}.$$

В результате получим, что расстояния

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n}$$

будут появляться с вероятностями

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}}{C_n^{n_\phi} C_n^{n_\psi}},$$

где $k = 0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$. Эти вероятности (и близкие к ним числа) можно использовать в качестве весов-коэффициентов расстояний для получения результирующего расстояния для данных формул с вероятностями (подробнее об этом см. дальше).

Заметим, что при таком подходе главную роль играют не сами формулы, а числа, определяющие количество моделей и их пересечения. Не имея другой информации, мы рассмотрели все возможные подмножества для подсчета частости (вероятности) появления расстояния для конкретного k. Используя свойство инвариантности расстояний между формулами и вероятностей высказываний (формул) [2–4], можно получить аналогичный результат, но с меньшим носителем знаний, включающим только те, которые встречаются в двух формулах, для которых и ищется расстояние. Будем считать, что мы так и сделали с самого начала. И тогда наш расчет будет оптимальным.

Зная вероятности p_k для каждого расстояния $\rho_k(B_i, B_j)$, в качестве расстояния между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ можно взять, например, наиболее вероятное расстояние $\rho(B_i, B_j) = \rho_m(B_i, B_j)$, где $p_m = \max_k p_k$, для которого справедлива теорема 1.

Для получения других расстояний можно использовать расстояние, усредненное для некоторых выбранных подмножеств из всех полученных расстояний.

Если мы возьмем произвольные веса p_k (исходя из экспертных оценок или дополнительных сведений экспертов, которые могут и не совпадать!) для расстояний $\rho_k(B_i, B_j)$ таким образом, чтобы они подчинялись закону распределения, то получим общий случай для адаптивного поиска нужного расстояния между формулами с вероятностями. Тогда в качестве расстояния между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ можно взять величину, равную математическому ожиданию (центру тяжести) или среднему значению случайной величины ξ:

$$\rho(B_i, B_j) = M\xi = \sum_{k=0}^{\min(n_\phi, n_\psi)} \rho_k(B_i, B_j) \cdot p_k.$$

Теорема 2. Для расстояния $\rho(B_i, B_j)$ справедливы свойства:

1) $0 \le \rho(B_i, B_j) \le 1$;

2) $\rho(B_i, B_j) = \rho(B_j, B_i)$;

3) $\rho(B_i, B_j) \le \rho(B_i, B_s) + \rho(B_s, B_j)$;

4) если $\rho(B_i, B_j) = 0$, то $B_i \equiv B_j$;

5) $\rho(B_i, B_j) = 1 - \rho(B_i, \neg B_j) = \rho(\neg B_i, \neg B_j)$;

6) $\rho(B_i, B_j) = \rho(B_i \wedge B_j, B_i \vee B_j)$.

Докажем неочевидное свойство 3. Для расстояния $B_i = \langle \phi, p_\phi \rangle$ по свойству 3 теоремы 1 имеем $\rho_k(B_i, B_j) \le \rho_k(B_i, B_s) + \rho_k(B_s, B_j)$. Тогда по свойствам математического ожидания:

а) $\xi \le \eta \Rightarrow M\xi \le M\eta$;

б) $M(\xi + \eta) = M\xi + M\eta$, откуда получаем требуемое свойство для расстояния.

Докажем свойство 4. Пусть $\xi(k) = \rho_k(B_i, B_j) \ge 0$ и $M\xi = 0$. Тогда по свойству математического ожидания $\xi(k) = \rho_k(B_i, B_j) = 0$ с вероятностью, равной 1, а по свойству 4 теоремы 1 $B_i \equiv B_j$.

Докажем свойство 5. Так как для расстояния $\rho_k(B_i, B_j)$ по свойству 6 теоремы 1 справедливо равенство

$$\rho_k(B_i, B_j) = 1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j),$$

тогда по свойствам математического ожидания:

а) если $p(\xi = \eta) = 1$ и $\exists M\xi$ (математическое ожидание существует), то $M\xi = M\eta$;

б) $M(a + b\xi) = a + bM\xi$, откуда получаем требуемое свойство для расстояния $\rho(B_i, B_j)$.

Таким образом, в данной статье исследованы способы введения метрик на классах эквивалентных высказываний экспертов, заданных формулами ИВ с вероятностями. Такое исследование необходимо для решения задач согласования вероятностных высказываний экспертов, кластеризации и для построения баз знаний и экспертных систем. Результаты переносятся на формулы над бесконечными носителям и формулы с переменными языка первого порядка с использованием измеримых (для фиксированной теории) подклассов измеримых (в том числе и метрических) моделей.

Библиографические ссылки

1. Блощицын В. Я., Лбов Г. С. О мерах информативности логических высказываний // Технология разработки экспертных систем : докл. Респ. школы-семинара. Кишинев, 1987. С. 12–14.
2. Лбов Г. С., Старцева Н. Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск : Изд-во Ин-та математики, 1999.
3. Vikent'ev A. A., Lbov G. S. Setting the metric and informativeness on statements of experts // Pattern Recognition and Image Analysis. 1997. Vol. 7 (2). P. 175–189.
4. Викентьев А. А., Лбов Г. С. О метризациях булевой алгебры предложений и информативности высказываний экспертов // Докл. Рос. акад. наук. 1998. Т. 361 (2). С. 174–176.
5. Бериков В. Б. Кластерный анализ с использованием коллектива деревьев решений // Науч. вестн. Новосиб. гос. техн. ун-та. 2009. № 3 (36). С. 67–76.
6. Лбов Г. С., Бериков В. Б. Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. Новосибирск : Изд-во Ин-та математики, 2005.
7. Ершов Ю. Л., Палютин Е. А. Математическая логика. М. : Наука,1991.

A. A. Vikentiev

INLETTING METRICS FOR EXPERT STATEMENTS WITH PROBABILITIES

In the work the author discusses ways of inletting metrics for expert statements represented as the formulas of Sentence Logic. Methods for inletting metrics on such formulas are offered and properties of the entered metrics are investigated. The research can be applied for solving problems of the best reconciliation of expert statements, for constructing decision functions in pattern recognition and for building expert systems.

Keywords: cluster analysis, expert statements, distance, metrics.