УДК 519.8+519.72

В. А. Лабковский

Об информативности наблюдения

Называя задачей решения (ЗР) упорядоченную тройку $Z = (\Theta, D, L)$, состоящую из произвольных множеств Θ, D и ограниченной действительной функции $L: \Theta \times D \to \mathbb{R}$, будем интерпретировать Θ, D и L как множество возможных значений неизвестного параметра, множество возможных решений и функцию потерь соответственно. Правилом выбора критерия (ПВК) будем называть отображение γ, сопоставляющее всякой задаче решения $Z = (\Theta, D, L)$ некоторую функцию $L_Z^*: D \to \mathbb{R}$ — критерий, имеющую смысл априорной оценки потерь: d_1 «лучше», чем d_2, если $L_Z^*(d_1) < L_Z^*(d_2)$.

Оказывается, если на используемые ПВК наложить некоторые естественные ограничения, то каждое ПВК будет порождаться каким-то семейством Ψ «конечно-аддитивных вероятностей» на Θ и каждое такое семейство будет порождать свое ПВК.

Интерпретируя Ψ как «информацию о θ», назовем пару (Z, Ψ) решающей системой (РС) и предположим, что прежде чем принимать решение, мы можем наблюдать значение известной функции $h: \Theta \to Y$. Как оценить информативность такого наблюдения? В настоящей работе предлагается один подход к решению этого вопроса. Приведен ряд теорем, показывающих естественность такого подхода. Данная работа продолжает исследования, начатые в [1, 2].

Оптимальным для данной РС наблюдением названо такое, которое имеет информативность не меньше, чем любое другое, и при этом мощность множества значений этого наблюдения не больше, чем у любого другого наблюдения с той же информативностью. Выясняется, что каждая РС имеет оптимальное наблюдение и каждое наблюдение оптимально для некоторой РС.

Переходя к точным формулировкам, удобно через Θ обозначить некоторое произвольное, но фиксированное множество, а через $\mathbb{Z}(\Theta)$ — класс всех упорядоченных троек вида $Z = (\Theta, D_Z, L_Z)$, где D_Z — какое-то множество, L_Z — ограниченная действительная функция, заданная на декартовом произведении $\Theta \times D_Z$.

О п р е д е л е н и е 1. *Правилом выбора критерия (ПВК) будем называть всякое отображение γ, заданное на $\mathbb{Z}(\Theta)$ и сопоставляющее каждой задаче решения $Z \in \mathbb{Z}(\Theta)$ действительную ограниченную функцию*

$$L_Z^*(\cdot) = \gamma(Z), \quad L_Z^*: D_Z \to \mathbb{R}.$$

При этом к классу Γ будем относить ПВК, удовлетворяющие следующим условиям:

1) *если* $Z_i = (\Theta, D_i, L_i)$, $d_i \in D_i$, $i = 1, 2$, $L_1(\theta, d_1) \leqslant L_2(\theta, d_2)$ $\forall \theta \in \Theta$, *то* $L_{Z_1}^*(d_1) \leqslant L_{Z_2}^*(d_2)$;

2) *если* $Z_i = (\Theta, D, L_i)$, $i = 1, 2$, $a, b \in \mathbb{R}$, $a \geqslant 0$, $L_1(\theta, d) = aL_2(\theta, d) + b$ $\forall \theta \in \Theta$, $d \in D$, *то* $L_{Z_1}^*(d) = aL_{Z_2}^*(d) + b$ $\forall d \in D$;

3) *если* $Z = (\Theta, D, L)$, $d_1, d_2, d_3 \in D$,

$$L(\theta, d_1) + L(\theta, d_2) = 2L(\theta, d_3) \quad \forall \theta \in \Theta, \tag{1}$$

то

$$L_Z^*(d_1) + L_Z^*(d_2) \geqslant 2L_Z^*(d_3). \tag{2}$$

Интуитивный смысл первых двух условий достаточно ясен, поэтому рассмотрим условие 3. Сравним две ЗР: $Z = (\Theta, D, L)$ и $\tilde{Z} = (\tilde{\Theta}, \tilde{D}, \tilde{L})$; где $\tilde{\Theta} = \Theta \times \Theta$, $\tilde{D} = D \times D$, $\tilde{L}((\theta_1, \theta_2), (d_1, d_2)) = L(\theta_1, d_1) + L(\theta_2, d_2)$. Интерпретируя $\tilde{Z}$ как двукратный выбор решения в одних и тех же условиях, естественно считать, что $L_{\tilde{Z}}^*(d_1, d_2) = L_Z^*(d_1) + L_Z^*(d_2)$, и поэтому неравенство (2) означает, что в предположении (1) лучше выбирать оба раза d_3, чем один раз d_1, а другой раз d_2. С чем это связано? Из (1) имеем

$$[L(\theta_1, d_3) + L(\theta_2, d_3)] - [L(\theta_1, d_1) + L(\theta_2, d_2)] =$$
$$= [L(\theta_2, d_1) + L(\theta_1, d_2)] - [L(\theta_1, d_3) + L(\theta_2, d_3)],$$

т. е. для любой пары значений θ_1 и θ_2 потери в $\tilde{Z}$ от применения (d_3, d_3) не зависят от порядка, в котором эти значения появляются, а если мы применим (d_1, d_2), то при одном порядке потери будут на столько же меньше, чем при (d_3, d_3),на сколько при другом — больше. Таким образом принятие условия 3 означает, что мы ориентируемся на гарантированный результат.

Дальше нам понадобятся еще некоторые обозначения. Пусть $PF(\Theta)$ — семейство всех конечно-аддитивных вероятностей на Θ:

$$PF(\Theta) = \{\psi \in (2^\Theta \to [0, 1]) : \psi(\Theta) = 1,$$
$$\psi(A \cup B) = \psi(A) + \psi(B \setminus A) \quad \forall A, B \subseteq \Theta\};$$

M — банахово пространство всех ограниченных действительных функций на Θ с нормой $\|f\| = \sup_{\theta \in \Theta} |f(\theta)|$; Q — множество всех конечных разбиений множества Θ на непересекающиеся подмножества; $\geqslant$ — естественное направление в Q:

$$Q_1 \geqslant Q_2 \Leftrightarrow \forall q_1 \in Q_1 \quad \exists q_2 \in Q_2, \quad q_1 \subseteq q_2.$$

Очевидно, для любых $\varepsilon > 0$, $f \in M$, $\psi \in PF(\Theta)$ можно указать такое $Q_\varepsilon \in \in Q$, что

$$\left| \sum_{q \in Q} f(\theta'_q)\, \psi(q) - \sum_{q \in Q} f(\theta''_q)\, \psi(q) \right| < \varepsilon$$

при $Q \geqslant Q_\varepsilon$, θ'_q, $\theta''_q \in q$, и поэтому предел интегральных сумм по направленному множеству $(Q, \geqslant)$ существует при любых $f \in M$, $\psi \in PF(\Theta)$. Обозначим его $\int f(\theta)\, \psi(d\widetilde{\theta})$.

Теорема. 1. *Всякому $\gamma \in \Gamma$ можно сопоставить такое множество $\Psi \subseteq PF(\Theta)$, $\Psi \neq \varnothing$, что при $Z \in \mathbb{Z}(\Theta)$, $L_Z^* = \gamma(Z)$ будет*

$$L_Z^*(d) = \sup_{\psi \in \Psi} \int L_Z(\theta, d)\, \psi(d\widetilde{\theta}) \quad \forall d \in D_Z. \tag{3}$$

Обратно, если $\Psi \subseteq PF(\Theta)$, $\Psi \neq \varnothing$, и отображение $\gamma : Z \mapsto L_Z^*$ на $\mathbb{Z}(\Theta)$ задается соотношением (3), *то* $\gamma \in \Gamma$.

Доказательство. Определим на M функционал $\varphi = \varphi_\gamma(\cdot)$ следующего вида. Всякой функции $f \in M$ сопоставим ЗР $Z_f = (\Theta, \{d\}, L_f)$, в которой $\{d\}$ — одноточечное множество, а $L_f(\theta, d) = f(\theta)$ при всех $\theta \in \Theta$. Теперь если $L_{Z_f}^* = \gamma(Z_f)$, то полагаем $\varphi(f) = L_{Z_f}^*(d)$. Очевидно, этот функционал не зависит от конкретного вида множества $\{d\}$ и, стало быть, он определен корректно.

Легко показать, что функционал φ обладает следующими свойствами:

1) если $f_1, f_2 \in M$, $f_1(\theta) \leqslant f_2(\theta)$ $\forall \theta \in \Theta$, то $\varphi(f_1) \leqslant \varphi(f_2)$;

2) если f_1, $f_2 \in M$, a, $b \in \mathbb{R}$, $a \geqslant 0$, $f_1(\theta) = af_2(\theta) + b$ $\forall \theta \in \Theta$, то $\varphi(f_1) = a\varphi(f_2) + b$;

3) если f_1, f_2, $f_3 \in M$, $f_1(\theta) + f_2(\theta) = 2f_3(\theta)$ $\forall \theta \in \Theta$, то $\varphi(f_1) + \varphi(f_2) \geqslant \geqslant 2\varphi(f_3)$.

Теперь обозначим через G множество всех ограниченных линейных функционалов g на M, удовлетворяющих двум условиям: 1) $g(1_\Theta) = 1$; 2) $g(f) \geqslant 0$, если $f(\theta) \geqslant 0$ $\forall \theta \in \Theta$.

Доказательство теоремы основывается на двух леммах, которые мы приведем без доказательств.

Лемма 1. *Между множествами $PF(\Theta)$ и G существует взаимно-однозначное соответствие $\psi \leftrightarrow g_\psi$, устанавливаемое соотношением*

$$g_\psi(f) = \int f(\theta)\, \psi(d\widetilde{\theta}) \quad \forall f \in M.$$

Лемма 2. *Пусть $G_\gamma = \{g \in G : g(f) \leqslant \varphi_\gamma(f)\ \forall f \in M\}$ при $\gamma \in \Gamma$. Тогда $G_\gamma \neq \varnothing$ и $\varphi_\gamma(f) = \sup_{g \in G_\gamma} g(f)$ $\forall f \in M$, $\gamma \in \Gamma$.*

Для завершения доказательства теоремы остается, пользуясь леммой 1, сопоставить всякому $\gamma \in \Gamma$ множество $\Psi_\gamma = \{\psi \in PF(\Theta) : g_\psi \in G_\gamma\}$ и на основании леммы 2 получить, что для любых $Z \in \mathbb{Z}(\Theta)$, $\gamma \in \Gamma$, $d \in D_Z$ при $L_Z^* = \gamma(Z)$ справедливы соотношения

$$L_Z^*(d) = \varphi_\gamma(L_Z(\cdot, d)) = \sup_{g \in G_\gamma} g(L_Z(\cdot, d)) = \sup_{\psi \in \Psi_\gamma} \int L_Z(\theta, d)\, \psi(d\widetilde{\theta}).$$

Тем самым первое утверждение теоремы доказано. Второе проверяется непосредственно.

Определение 2. *Решающей системой (РС) будем называть всякую упорядоченную четверку вида $S = (\Theta, D, L, \Psi)$, где $(\Theta, D, L) \in \mathbb{Z}(\Theta)$, а Ψ — произвольное непустое подмножество множества $PF(\Theta)$. Класс всех РС будем обозначать через $\mathrm{S}(\Theta)$ и сопоставлять всякой РС $S = (\Theta, D,$*

$L, \Psi)$ *функцию* $L_S^*: D \to \mathbb{R}$, $L_S^*(d) = \sup\limits_{\psi \in \Psi} \int L(\theta, d)\, \psi(d\theta)$, *а также число* $\rho(S) = \inf\limits_{d \in D} L_S^*(d)$.

Определение 3. *Наблюдением на Θ будем называть всякую функцию (не обязательно числовую) с областью определения Θ. Класс всех наблюдений на Θ будем обозначать через $H(\Theta)$ и называть информативностью наблюдения h относительно PC $S = (\Theta, D, L, \Psi)$ величину $INF(h/S) = \rho(S) - \rho(S^h)$, где S^h — новая PC, определяемая следующим образом:*

$$\Delta = (h(\Theta) \to D),\ \mathcal{L}: \Theta \times \Delta \to \mathbb{R},\ \mathcal{L}(\theta, \delta) = L(\theta, \delta(h(\theta))), \quad S^h = (\Theta, \Delta, \mathcal{L}, \Psi),$$

$h(\Theta)$ — *образ множества Θ при отображении h.*

Наконец будем говорить, что наблюдение $h_1 \in H(\Theta)$ тоньше, чем $h_2 \in H(\Theta)$, когда для любых θ_1 $\theta_2 \in \Theta$ из $h_1(\theta_1) = h_1(\theta_2)$ следует $h_2(\theta_1) = h_2(\theta_2)$.

Теорема 2. *Наблюдение $h_1 \in H(\Theta)$ тоньше, чем наблюдение $h_2 \in H(\Theta)$, тогда и только тогда, когда*

$$INF(h_1/S) \geqslant INF(h_2/S) \quad \forall S \in \mathbf{S}(\Theta).$$

Доказательство. Пусть h_1 не тоньше, чем h_2, т. е. существуют такие $\theta_0, \theta_1 \in \Theta$, что $h_1(\theta_0) = h_1(\theta_1)$ и в то же время $h_2(\theta_0) \neq h_2(\theta_1)$. Построим PC следующего вида: $D = \{d_0, d_1\}$ — двухточечное множество,

$$L: \Theta \times D \to \mathbb{R}, \quad L(\theta, d_i) = \begin{cases} 1, \text{ если } \theta = \theta_i, \\ 0, \text{ если } \theta \neq \theta_i, \end{cases} \quad i = 0, 1, \quad \Psi = PF(\Theta),$$

$$S = (\Theta, D, L, \Psi).$$

Обозначим $S^{h_i} = (\Theta, \Delta_i, \mathcal{L}_i, \Psi)$, $i = 1, 2$. Непосредственной проверкой легко убедиться в том, что $\rho(S) = \inf\limits_{d \in D} \sup\limits_{\theta \in \Theta} L(\theta, d) = 1$. Кроме того, поскольку $h_1(\theta_0) = h_1(\theta_1)$, ясно, что $\delta_1(h_1(\theta_0)) = \delta_1(h_1(\theta_1))$ при всех $\delta_1 \in \Delta_1$, и значит, $\sup\limits_{\theta \in \Theta} L(\theta, \delta_1(h_1(\theta))) = 1$ $\forall \delta_1 \in \Delta_1$. Таким образом,

$$\rho(S^{h_1}) = \inf_{\delta_1 \in \Delta_1} \sup_{\theta \in \Theta} L(\theta, \delta_1(h_1(\theta))) = 1.$$

С другой стороны, взяв $\delta_2 \in \Delta_2$ вида $\delta_2(y) = \begin{cases} d_1, \text{ если } h_2(\theta_0) = y, \\ d_0, \text{ если } h_2(\theta_0) \neq y, \end{cases}$ получим $L(\theta, \delta_2(h_2(\theta))) = 0$ $\forall \theta$, и значит, $\rho(S^{h_2}) = 0 < \rho(S^{h_1}) = 1$. Прямое утверждение теоремы доказано. Обратное утверждение доказывается несложно, поскольку для любого $\delta_2 \in \Delta_2$ легко указать такое $\delta_1 \in \Delta_1$, что $\delta_2(h_2(\theta)) = \delta_1(h_1(\theta))$ $\forall \theta \in \Theta$.

Обозначим

$$H_* = \{h \in H(\Theta): h(\theta_1) = h(\theta_2) \ \forall \theta_1, \theta_2 \in \Theta\},$$

$$H^* = \{h \in H(\Theta): h(\theta_1) \neq h(\theta_2), \text{ если } \theta_1 \neq \theta_2\}.$$

Теорема 3. *Пусть $h_* \in H_*$, $h^* \in H^*$, $h \in H(\Theta)$. Тогда*

$$0 = INF(h_*/S) \leqslant INF(h/S) \leqslant INF(h^*/S) \quad \forall S \in \mathbf{S}(\Theta).$$

Доказательство. Очевидно, h^* тоньше, чем h, и h тоньше, чем h_*, поэтому надо лишь доказать, что $INF(h_*/S) = 0$, т. е. $\rho(S^{h_*}) = \rho(S)$. Для этого заметим, что множество $h_*(\Theta)$ состоит из одной единственной точки и, следовательно,

$$\rho(S^{h_*}) = \inf_{\xi \in \Delta_{S^{h_*}}} \sup_{\psi \in \Psi} \int L(\theta, \delta(h_*(\theta)))\, \psi(d\theta) = \inf_{d \in D} \sup_{\psi \in \Psi} \int L(\theta, d)\, \psi(d\theta) = \rho(S),$$

что и требовалось доказать.

Определение 4. *Наблюдение $h \in H(\Theta)$ будем называть оптимальным для РС $S \in \mathbf{S}(\Theta)$ тогда и только тогда, когда $INF(h/S) \geqslant INF(h_1/S)$ $\forall h_1 \in H(\Theta)$, и если $INF(h/S) = INF(h_1/S)$, то* $\mathrm{Card}\,(h(\Theta)) \leqslant \mathrm{Card}\,(h_1(\Theta))$, *где* $\mathrm{Card}\,(A)$ *— мощность множества A.*

Теорема 4. *Для любой РС можно указать оптимальное наблюдение и любое наблюдение оптимально для некоторой РС.*

Доказательство. Обозначим через $\mathbf{H}(S)$ класс наблюдений, имеющих относительно PC $S \in \mathbf{S}(\Theta)$ наибольшую информативность, т. е. информативность, равную, согласно теореме 3, $\nu(S) = INF(h^*/S)$. Каждому $h \in \mathbf{H}(S)$ соответствует некоторая мощность множества $h(\Theta)$, и мы можем рассмотреть множество всех таких мощностей: $\{\mathrm{Card}\,(h(\Theta)) \mid h \in \mathbf{H}(S)\}$. Как известно (см. [3, с. 85]), всякое множество мощностей является вполне упорядоченным по величине, поэтому среди наблюдений $h \in \mathbf{H}(S)$ найдется такое, у которого мощность множества $h(\Theta)$ минимальна. Очевидно, оно и будет оптимальным для S.

Обратно, пусть $h \in H(\Theta)$. Положим $D = h(\Theta)$,

$$L(\theta, d) = \begin{cases} 0, & \text{если } h(\theta) = d, \\ 1, & \text{если } h(\theta) \neq d; \end{cases}$$

$$\Psi = PF(\Theta), \quad S = (\Theta, D, L, \Psi).$$

Очевидно, для такой РС справедливы равенства

$$\rho(S) = \inf_{d \in D} \sup_{\theta \in \Theta} L(\theta, d) = 1, \quad \rho(S^h) = \inf_{\delta \in \Delta} \sup_{\theta \in \Theta} L(\theta, \delta(h(\theta))),$$

и если δ_0 — тождественное преобразование множества D, то

$$\rho(S^h) \leqslant \sup_{\theta \in \Theta} L(\theta, \delta_0(h(\theta))) = 0.$$

С другой стороны, очевидно, что $\rho(S^h) \geqslant \inf_{\theta, d} L(\theta, d) = 0$, и таким образом получаем

$$\rho(S^h) = 0, \quad INF(h/S) = 1.$$

Легко проверяется, что и $\nu(S) = 1$, и значит, остается лишь доказать что если $h_1 \in \mathbf{H}(S)$, то

$$\mathrm{Card}\,(h_1(\Theta)) \geqslant \mathrm{Card}\,(h(\Theta)).$$

Пусть $h_1 \in \mathbf{H}(S)$. Тогда для любого $\varepsilon > 0$ можно указать такое $\delta_1 \in (h_1(\Theta) \to D)$, что $\sup_{\theta \in \Theta} L(\theta, \delta_1(h_1(\theta)) < \varepsilon$, и поскольку функция L принимает значения только 0 или 1, это означает, что $L(\theta, \delta_1(h_1(\theta))) = 0$, т. е. $\delta_1(h_1(\theta)) = h(\theta)$ $\forall \theta \in \Theta$. Иными словами, существует отображение δ_1 множества $h_1(\Theta)$ *на* множество $h(\Theta)$, а это возможно лишь при условии, что мощность множества $h(\Theta)$ не превышает мощности множества $h_1(\Theta)$.

Теорема 5. *Пусть $S = (\Theta, D, L, \Psi)$, $h \in H(\Theta)$ — наблюдение, оптимальное для S, и множество D конечно. Тогда*

$$\mathrm{Card}\,(h(\Theta)) \leqslant \mathrm{Card}\,(D).$$

Доказательство. Поскольку множество D конечно, можно сопоставить всякому $\theta \in \Theta$ непустое подмножество множества D, на котором функция $L(\theta, \cdot)$ принимает минимальное значение $D_\theta = \operatorname{Argmin}_{d \in D} L(\theta, d)$. Пусть $\bigcup_{\theta \in \Theta} D_\theta = D^0 = \{d_1, d_2, \ldots, d_n\}$. Обозначим $\Theta_1 = \{\theta \in \Theta : D_\theta \ni d_1\}$, $\Theta_i = \left\{\theta \in \Theta : D_\theta \ni d_i,\ \theta \notin \bigcup_{j=1}^{i-1} \Theta_j\right\}$, $i = 2, 3, \ldots, n$. Очевидно, $\bigcup_{i=1}^{n} \Theta_i = \Theta$, $\Theta_i \cap \Theta_j = \varnothing$ при $i \neq j$. Положим $h_1 : \Theta \to \{1, 2, \ldots n\}$, $h_1(\theta) = i$ при $\theta \in \Theta_i$. Будем иметь

$$\rho(S^{h_1}) = \inf_{\delta \in \Delta} \sup_{\psi \in \Psi} \int L(\theta, \delta(h_1(\theta)))\, \psi(\widetilde{d\theta}) = \sup_{\psi \in \Psi} \int [\min_{d \in D} L(\theta, d)]\, \psi(\widetilde{d\theta}),$$

откуда легко следует, что $INF(h_1/S) = \nu(S)$, и если какое-то наблюдение h оптимально для S, то

$$\mathrm{Card}\,(h\,(\Theta)) \leqslant \mathrm{Card}\,(h_1\,(\Theta)) \leqslant n \leqslant \mathrm{Card}\,(D).$$

1. *Иваненко В. И., Лабковский В. А.* О функции неопределенности байесовских систем.— Докл. АН СССР, 1979, **248**, № 2, с. 307—309.
2. *Лабковский В. А.* Об информативности стохастического эксперимента.— Кибернетика, 1982, № 4, с. 90—93.
3. *Александров П. С.* Введение в теорию множеств и общую топологию.— М. : Наука, 1977.— 366 с.

Ин-т кибернетики АН УССР, Киев

Получено 11.05.85,
после доработки — 18.09.85