О ФОРМАЛИЗАЦИИ ПРОЦЕССОВ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА

УДК 6.21.377.037

СОКОЛОВ А.Е.
Докторант
Херсонский национальный технический университет, г. Херсон, Украина
Научные интересы: компьютеризованные системы обучения

ВВЕДЕНИЕ

Массовое использование компьютерной техники и информационных технологий обусловило большой интерес к вопросам создания информационных систем, принципам их построения. Представляют также интерес вопросы формализации процессов информационных систем, в том числе и информационного пространства. Под информационным пространством чаще всего понимают совокупность:

– баз и банков данных;

– технологий их применения;

– информационных коммуникационных систем, которые функционируют

– на базе общих принципов и обеспечивают информационное взаимодействие объектов разной природы, а также взаимодействие человека, организационных систем и решают задачи удовлетворения информационных потребностей пользователей.

Целью данной работы является решение и систематизация процессов формализации отдельных задач информационного пространства решении прикладных задач информационных технологий. Отдельные задачи формализации процессов в информационных системах, в том числе и определения количества информации изложены в работах [1-4]. Не смотря на предложенные многочисленные методы задач эти ещё далеки от решения этой проблемы.

ОСНОВНОЕ СОДЕРЖАНИЕ

Основой понятия пространства является множество, наделяемое структурой измерения величины элемента – нормой и измерения расстояния между элементами – метрикой [1].

Традиционно множеством, в теории в информации, является множество событий, над которым определяются операции над элементами и алгебра событий [5]. Полученное таким образом пространство событий наделяется вероятностью – нормой в вероятностном пространстве [6]. Более сложно обстоит вопрос с измерением расстояния между событиями. Наиболее естественно использовать в качестве метрики условные вероятности, что определяет статистическую близость событий [3].

При этом необходимо учесть, что вероятность события величина измеримая и существует простой рецепт определения вероятности наблюдаемых событий как отношения количества событий а к общему количеству событий

$$P_a = \frac{n_a}{n}. \quad (1)$$

Таким образом для пространства событий A, с алгеброй событий ξ получаем вероятностное пространство E с нормой [7]

$$\|E\| = P_a,. \quad (2)$$

и метрикой [3]

$$a(a_k, a_m) = P(a_k/a_k) - P(a_k/a_m). \quad (3)$$

Естественно алгебра событий ξ, в вероятностном пространстве порождает алгебру теории вероятностей ξe [7]. Однако возможность определения свойств объектов с их вероятностной стороны недостаточна для описания процессов со стороны причин явлений, явно определяющих события материального мира. Необходимость введения новой величины, определяющей причину события, первыми ощутили физики [8], что определило появление понятия энтропии.

В дальнейшем развитии появилось понятие энтропии, применительно к задачам передачи информации [4] и наконец сформировалось понятие информации, определяемое через вероятность события (1)

$$I_a = -\log_c P_a. \quad (4)$$

Со временем пришло осознание энтропии как ожидаемой информации (4)

$$H_a = M\{I_a\}. \quad (5)$$

Собственно следующий шаг [9] логически привел к информационному пространству, как продолжению вероятностного пространства (1, 2, 3) с нормой

$$\|I_a\| = -\log_c P_a.$$

И метрикой

$$a(I_k, I_m) = -\log_c P(a_k/a_m). \quad (6)$$

Естественно наследование алгебры и всего развития теории вероятностей.

Однако существует противоречие – процессы в природе не ограничены стохастическими, а развитый, на сегодняшний день, механизм теории информации ограничен метрикой (5) и нормой (6).

Данное противоречие снимается очень просто, достаточно рассмотреть (4) и отметить, что по сути дела разговор идет о косвенных измерениях и существует прямая зависимость

$$I = I(P) \Leftrightarrow P = P(I). \quad (7)$$

Следовательно для измерения информации на основе измеренной вероятности исходным отношением является зависимость вероятности от информации (7). Таким образом все становится на свои места и не информация зависит от вероятности, а вероятность события зависит от информации [9]. Неожиданно простое утверждение – информация это причина события резко упрощает картину, действительно достаточно предположить аналитичность вероятности, как мы получаем разложение вряд

$$P = P_0 + \frac{1}{1!}\frac{dP}{dI_{|I=I^*}}\Delta I + \frac{1}{2!}\frac{d^2P}{dl^2_{|I=I^*}}\Delta I^2 + \ldots R\,. \quad (8)$$

где: P0 – вероятность до прихода нового сообщения;

I* – информация в ранее принятом сообщении; ΔI приращение информации с новым сообщением.

Таким образом количество информации ΔI, вызвавшей изменение вероятности ΔP, можно найти обращая (8).

$$\Delta I = P^{-1}(\Delta P). \quad (9)$$

Собственно это стандартная ситуация возникающая при моделировании – учет не только линейных членов разложения при обращении ведет к повышению точности, но при малых изменениях основную роль играет линейная часть разложения, что ведет к линеаризации. В таком случае, для P0 = 0, и ограничившись линейным приближением (9) получаем

$$P \approx \frac{1}{1!}\frac{dP}{dI}\Delta I. \quad (10)$$

Собственно в (10) производная вероятности по информации это просто плотность вероятности f(I) и мы можем записать

$$\frac{dP}{dI} = f(I). \quad (11)$$

Следовательно можем, зная f(I), просто найти связь между вероятностью и информацией

$$dP = f(I)dl. \quad (12)$$

Далее решая полученное дифференциальное уравнение (12)

$$\int dP = \int f(I)dl. \quad (13)$$

Из (13) находим P=P(1) обращая решение как полученную композицию I=P-1(P).

Таким образом получаем возможность определить норму в информационном пространстве, однако остается вопрос о виде плотности f(I).

Здесь необходимо найти плотность распределения f(I)=-αP(I). Для этого необходимы исследования процесса, проверка гипотезы о распределении [9] после чего получаем дифференциальное уравнение для данного процесса. Например, если предположить, что процесс подчиняется уравнению органического роста

$$\frac{dP}{dI} = -\alpha P,$$

то получаем известную меру Хартли

$$\int \frac{dP}{P} = -\alpha \int dI, \rightarrow\ I = -\frac{1}{a} \ln P.$$

Собственно далее все просто – гипотеза f(I)=-αP(I) порождает информационное пространство с нормой (5) и метрикой (6) где энтропия определяется ожидаемой информацией.

Однако здесь начинается самое интересное, так как функция распределения определяется процессом, придется в каждом конкретном случае определяться с функцией распределения. Секрет удачности меры Хартли в том, что для канала связи действительно скорость изменения вероятности пропорциональна текущей вероятности события. Одновременно этим объясняются трудности развития теории информационных систем, где рассматриваются различные задачи, ожидания, сбора информации, хранения информации и так далее, но мера Хартли адекватна только задаче передачи информации. Как следствие множество попыток найти эмпирические меры [10].

Таким образом, необходимо рассмотрение типовых процессов, описываемых в информационном пространстве. Так, для простейшего случая несвязности информации и события, имеем

$$\frac{dP}{dI} = 0.$$

Следовательно в этом случае P=Const и P≠P(I).

В случае ожидания события [9], например приход синхросигнала

$$\frac{dP}{dl} = \alpha, \rightarrow I = \frac{1}{\alpha} P. \quad (14)$$

Действительно если мы используем меру Хартли, то собственно ожидаемый сигнал не информативен, с другой стороны практика говорит о обратном, мы формируем окна, строим фильтры именно для обнаружения ожидаемого сигнала, стремясь снизить влияние маловероятных сигналов – помех.

Другой, практически важный случай, связан с процессом сбора информации от множества источников [9]. В этом случае для множества событий Ω, каждое из которых генерируется отдельным независимым источником, в силу центральной предельной теоремы, имеем к нормальное распределение

$$\frac{dP}{dI} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(I-ml)^2}{2\sigma^2}}. \quad (15)$$

где: ml – ожидаемая информация, σ – среднеквадратическое отклонение.

Собственно данное утверждение это только гипотеза и, естественно, все дальнейшее правильно, если гипотеза верна, но гипотеза вполне оправдана и реализуема. Исходя из связи (15) для получения явной оценки информации по измеряемой величине – вероятности необходимо решить дифференциальное уравнение

$$dP = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(I-ml)^2}{2\sigma^2}} dl. \quad (16)$$

Естественно из (16) получаем функцию Лапласа, как связь между вероятностью и информацией, в случае нормального распределения

$$P = \frac{1}{\sqrt{2\pi\sigma}} \int e^{-\frac{(I-ml)^2}{2\sigma^2}} dl\,; mI = 0, \sigma = 1, \quad (17)$$

или, для принятых значений математического ожидания и среднеквадратического отклонения, получаем

$$\Phi(I_x) = \frac{1}{\sqrt{2\pi}} \int_0^{I_x} e^{-\frac{I^2}{2\sigma^2}} e^{-\frac{(I-ml)^2}{2}} dI = P(I_x).$$

Таким образом, нормой в информационном пространстве, в рассматриваемом случае сбора информации, является доверительный интервал для I с доверительной информацией Px, используя функцию Лапласа можем записать (17) в виде

$$\|I\| = \Phi^{-1}(P_x, I_x), \quad (18)$$

Следовательно для интервала (α, β) имеем

$$\|\Delta I\| = \left[\Phi\left(P_x, \frac{I_\beta - mI}{\sigma}\right) - \Phi\left(P_x, \frac{I_\alpha - mI}{\sigma}\right)\right]^{-1}.$$

Таким образом, величина информации в событии определяется доверительной вероятностью нахождения информации I в заданном интервале $(I_\alpha,\ I_\beta)$ для этого события, и чем больше мы можем доверять полученному результату, тем выше его информативность. Естественно, в этом случае справедливо правило сигм.

Обозначив математическое ожидание и среднеквадратическое отклонение в (18) как

$$m^*I = \rho_{II^*}I^*, \sigma^* = \sqrt{1-\rho_{II^*}^2}.$$

Получаем метрику как доверительный интервал для отклонения

$$a(I,I^*) = \left[\frac{1}{\sigma_I}\left(\Phi\left(P_{I^*},\frac{I_\beta - m^*I}{\sigma^*}\right) - \Phi\left(P_{I^*},\frac{I_\alpha - m^*I}{\sigma^*}\right)\right)\right]^{-1}.$$

При этом выполняются требования к метрике, не отрицательность, аксиома тождества выполняется при полной уверенности Px=1, аксиома треугольника. Однако аксиома симметрии, в данном случае, не применима.

Важным свойством информационного пространства является возможность построения его не только над вероятностным пространством, но и над детерминированными процессами. Так в работе [11], построено информационное пространство в задаче оптимального хранения информации.

В этом случае информация связана с достижением определенной цели, заданной конкретной целевой функцией

$$f_i = f_i(I_i) \tag{19}$$

Далее повторяя разложение (8) для целевой функции (19) получаем

$$f(I) = f(I_0) + \frac{1}{1!}\frac{df}{dI_{I=I_0}}\,\Delta I + \frac{1}{2!}\frac{d^2f}{dI^2_{I=I_0}}\Delta I^2 \ldots + R \tag{20}$$

где: I0 – наличная информация,

Рассмотрим в (20) линейное приближение и предположим, что скорость изменения функции цели пропорциональна наличной информации, что характерно для задачи обучения – мы тем быстрее учимся, чем мы умнее. Тогда можно записать

$$\frac{df}{dI} = \alpha I. \tag{21}$$

Следовательно из (21) получаем

$$f = \frac{\alpha}{2}I^2.. \tag{22}$$

Используя в (22) граничные условия f(Im)=fm, определим коэффициент α

$$\alpha = \frac{2f_m}{I_m^2}.$$

Тогда связь между функцией цели и информацией в данной задаче принимает вид

$$\|I\| = I_m\sqrt{\frac{f}{f_m}}. \tag{23}$$

Таким образом, выражение (23) определяет норму в информационном пространстве для рассматриваемого случая.

Метрика в данной задаче, будет определяться условным изменением целевой функции

$$a(I_x, I_y = I_m\sqrt{\frac{f_{x/y}}{f_m}}. \tag{24}$$

Так как в (24), $f_{x/y}$ неотрицательна, расстояние всегда реально и неотрицательно, это обеспечивает выполнение аксиомы неотрицательности.

Таким образом, норма в информационном пространстве определяется особенностью процесса и различна для различных процессов. В таблице 1 сгруппированы рассмотренные процессы и их нормы в информационном пространстве.

Таблица 1.

Нормы в информационном пространстве для различных процессов

Естественно данные таблицы 1 не охватывают всего множества информационных процессов, но показывают различие норм информационного пространства для различных процессов.

ВЫВОДЫ

1. Определение информации, как причины события, позволило в задачах формализации рассматривать нормированное, метрическое информационное пространство, где процессы формализации информационных систем получают единый механизм анализа и синтеза.

2. Единство подхода к определению нормы и метрики в информационном пространстве создает основу для дальнейшего развития теории информации.

3. Сложность и многообразие информационных процессов выдвигают требование определения норма и метрики для каждого конкретного процесса, что естественно усложняет процесс формализации, но обеспечивает адекватность методов расчета.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1. T.B.Ward Functional analysis lecture notes / T.B.Ward School of Mathematics, University of East Anglia, 2000.
2. Kolmogorov A.H. Teorija informacii i teorija algoritmov / A.N. Kolmogorov -M.: «Nauka», 1987. -304 s.
3. Stratanovich R.L. Teorija informacii / Stratanovich R.L. – M.: Sov. Radio, 1975. – 424 s.
4. Hartley R. V. L. Transmission of information / Hartley R. V. L. // Bell System Technical Journal – 7. – 1928. – S. 535 -563.
5. Kolde Ja.K. Praktikum po teorii verojatnostej i matematicheskoj statistike / Kolde Ja.K – M.: «vysshaja shkola», 1991. -157 s.
6. Jaglom A.M. Verojatnost' i informacija / A.M.Jaglom, I.M.Jaglom – Izd. 5-e, stereotipnoe.-M.: KomKniga, 2007. – 512 s.
7. M. Kendell Teorija raspredelenij / Kendell M., St'juart, pod. Red. A. N. Kolmogorova. – T1, M.: «Nauka», 1966. – 573 s.
8. Je.Fermi Termodinamika. / Je.Fermi. – Izhevsk: RHD, 1998.-164 s.
9. Brazhnik D. O. Modeli ta metodi pidvishhennja stijkosti do pereshkod v optichnih sistemh identifikacії: dis... kandidata teh. nauk: 05.13.06 / Brazhnik D. O. – Herson. 2012 .- 170 s.
10. Dzh. Tu, R.Gonsales Principy raspoznavanija obrazov / Dzh. Tu, R.Gonsales/- M.: «Mir», 1978. – 411 s.
11. Sokolova O. V. Modeli i metody informacionnyh tehnologij nakoplenija informacii: dis... kandidata teh. nauk: 05.13.06 / Sokolova O. V. – Herson. 2014. 190 s.

Рецензент: д.т.н., проф. Ходаков В.Е.
Херсонский национальный технический университет