

КЛАССИЧЕСКАЯ ТЕОРИЯ ИНФОРМАЦИИ БЕЗ ОГРАНИЧЕНИЙ

Пантюшин В.А.,

К.т.н, доцент, Государственный университет по землеустройству (г. Москва)

geotopic@mail.ru

Аннотация. В статье на основе критического анализа различных подходов и взглядов на практические приложения теории информации обосновывается необходимость в разработке новых алгоритмов количественной оценки информации, предполагающих учет не только статистических связей элементов, но и разнопараметрические оценки этих связей с учетом системного анализа.

Ключевые слова: количество информации, теория вероятности, математическая теория связи, передача и обработка данных, критерий информативности, информативный образ, дешифрирование аэрокосмических изображений.

THE CLASSICAL THEORY OF INFORMATION WITHOUT RESTRICTIONS

Pantyushin V.A.

State University Of Land Use Planning (Moscow)

Abstract. In the article on basis of review of the different approaches and views on practical applications of information theory is grounded on the necessity for elaboration new algorithms of quantitative assessment of information, taking into account not only the statistical relationships of elements, but also different parametric assessments of these relationships based on systems analysis.

Keywords: amount of information, theory of probability, mathematical theory, communication theory, transmission and handling data, informative qualities, sacred image, interpretation aerospace representation.

Разработанные Хартли Р. и Шеноном К. методы обработки и оценки сообщений, получили широкое применение при передаче информации по техническим каналам связи. Разработанная ими теория информации позволила получить количественную оценку информационных сообщений, дала возможность определять и рассчитывать необходимые параметры передающих средств в зависимости от объема информации, и положила начало новым методам обработки сообщений в криптографии и криптоанализе, кодированию и сжатию информации. Находит активное применение теория информации в дистанционном зондировании, фотограмметрии и дешифрировании аэрокосмических изображений. Неслучайно термин «дешифрирование» (от франц. déchiffrer - разбирать, разгадывать) имеет ту же аналогию в теории связи, криптографии и дистанционном зондировании. До недавнего времени, эффективное использование методов «математической теории связи» в дешифри-

ровании аэрокосмических изображений, в большей степени, сдерживалось возможностями технических средств по реализации алгоритмов количественной оценки несжатой цифровой информации об объектах Земной поверхности различного разрешения.

Необходимо отметить, что разработка «шеноновской теории информации» по времени совпала со стремительным внедрением вычислительной техники в различные сферы деятельности и отрасли производства, с развитием кибернетики, теории автоматов и вычислительных машин. То есть, период важных открытий в области математической теории связи совпал по времени с началом эры информационного общества. Все это вызвало стремление перенести новый математический аппарат теории вероятностей и математической статистики, положенный в основу «новой теории информации» на решение частных задач разных отраслей знания и практики в целях достижения конкретных результатов и фактов. «Теория информации, как мод-

ный опьяняющий напиток, кружит голову всем вокруг» /1/ и до настоящего времени. В своих работах Шенон отмечал, что «поиск путей применения теории информации в других областях не сводится к тривиальному переносу терминов из одной области науки в другую», а требует глубокого понимания математической стороны теории информации и ее практических приложений, длительного процесса выдвижения новых гипотез и их экспериментальной проверки. Кроме того, «основные положения теории информации касаются очень специфического направления исследования, направления, которое совершенно не обязательно должно оказаться плодотворным в психологии, экономике и в других социальных науках» /1/.

Наличие различных подходов и взглядов на практические приложения теории информации, а так же существующие споры об ограничениях классической теории информации, обусловлены не только «отсутствием в теории информации дефиниции (определения) самой информации» /3/, но в большей степени попытками перенести математический аппарат теории информации, разработанный для приложения к теории связи периода 40-х годов прошлого века, на технологии передачи и обработки данных (информации) современными системами и мозгом человека. Примером такого «тривиального переноса» математического аппарата «шенноновской теории информации» на смежные области науки и современные технологии могут служить так называемые «парадоксы теории информации», рассмотренные авторами в ряде публикаций и научных трудах /2/, /3/.

Так в рассматриваемом «парадоксе с перестановкой букв» /2/ отмечается, что в классической теории информации слово “буква” несет столько же информации, сколько и слово “укваб” и количество информации, содержащееся в данном слове (сообщении, состоящем из М букв), может быть вычислено по формуле /1/ (1):

$$B_m = -M \sum_{j=1}^n P_j \log_2 P_j, \quad (1)$$

где n - число букв в данном языке, а p_j - частота встречаемости j -той буквы ($j = 1, 2, \dots, n$) в этом языке.

По мнению автора «самостоятельное слово “буква” не несет никакой информации без контекста» и «нести информацию могут лишь определенные сочетания букв, ...отдельные буквы, или поставленные как попало буквы, не несут информации» /2/. То есть, использование формулы (1) для оценки информации в слове без учета «контекста» не имеет смысла.

Однако напрашивается вопрос, а что мешает использовать эту же формулу для оценки информации в сочетании букв (слов) с учетом контекста? При таком подходе достаточно под p_j рассмотреть частоту встречаемости j -той биграммы, триграммы и более сочетания букв (слов). И такие исследования также проведены. Нижеприведенный текст с перестановкой букв, свидетельствует о том, что семантическая нагрузка появляется от определенной информативной последовательности букв или словосочетаний, трансформирующихся в сознании в определенные словесные образы (рис.1).

По результатам исследования одного английского университета, не имеет значения, в каком порядке расположены буквы в слове. Голубое, чтобы прева и последняя буквы были на месте. Остальные буквы читаются без проблем. Причиной этого является то, что мы не читаем каждую букву по отдельности, а все слово целиком.

Рис. 1. Пример «парадокса с перестановкой букв».

Причем, для образного восприятия даже не важна последовательность букв в словосочетании: “буква” и “буква” действительно несут одинаковое количество информации и в контексте всего сообщения, при желании, абстрагируясь от грамматических требований, учувствуют в формировании словесных образов от прочитанной информации.

Повысить точность оценки информации в сообщении, поможет и выявление наиболее информативных сочетаний групп букв применительно для каждого языка. Критерий информативности требует особого обоснования. Так в нижеследующем фрагменте сообщения (рис. 2), именно информативные комбинации групп букв позволяют прочитать, на первый взгляд, не имеющий ни какого смысла текст. Становится действительно интересно, какие удивительные вещи может различать наш разум!¹

**94НН03 С006ЩЗНН3 ПОК4ЗЫ8437, К4КИЗ
У9И8И73ЛЬНЫЗ 8ЗЩИ МОЖЗ7 9ЗЛ47Ь
Н4Ш Р4ЗУМ! 8ПЗЧ47ЛЯЮЩИЗ 8ЗЩИ!
СР4Ч4Л4 Э7О БЫЛО 7РУ9НО, НО СЗЙЧ4С
Н4 Э7ОЙ С7РОКЗ 84Ш Р4ЗУМ ЧИ7437 Э7О
487ОМ47ИЧЗСКИ, НЗ 349УМЫ84ЯСЬ О6
Э7ОМ. ГОР9ИСЬ! ЛНШЬ ОПР39ЗЛЗННЫЗ
ЛЮ9Н МОГУ7 ПРОЧН747Ь Э7О**

Рис. 2. Пример информативных комбинаций букв, определяющих смысл сообщения.

Парадокс с расчетом по типовой формуле количества “принятой глазом информации” /2/, /4/ обусловлен уровнем знания в конкретной области практического приложения теории информации, а так же уровнем развития вычислительной техники. Так, по состоянию на 40-60 годы прошлого века, использование формулы Хартли (2) для оценки информации в изображении черно-белого негатива весьма затруднительно, так как возникали сложности вычисления степенной функции в зависимости от числа анализируе-

¹ На данном фрагменте текста содержится следующее сообщение: «Данное сообщение показывает какие удивительные вещи может различать наш разум! Впечатляющие вещи! Сначала это было трудно, но сейчас на этой строке наш разум читает это автоматически, не задумываясь об этом! Гордись! Лишь определенные люди могут прочитать это».

мых элементов изображения даже при 256 значениях градаций яркости:

$$B = \log_2 M^N, \quad (2)$$

где N - число элементов изображения, M - число градаций яркости.

Еще большую сложность, по мнению авторов /2/, вызывает вычисление по формуле (2) количества «принятой глазом информации». Так как «число градаций яркости M глаза остается неопределенным, то и вычисления по формуле (2) можно считать достаточно случайными».

Современное состояние знаний в области медицины, теории зрения, дистанционного зондирования, фотограмметрии и геоинформационных систем позволяют достаточно точно определять и использовать для вычисления объемов информации цифровых изображений не только число градаций яркости черно-белых негативов, но и цветных. Так же достаточно точно определено, на современном уровне, не только число градаций яркости M, но и спектральное разрешение человеческого глаза.

Спектральное разрешение зрения человека связано со способностью глаза различать характеристики цвета: цветовой тон (λ), чистоту цвета (ρ) и яркость (B).

Порог цветоразличения по цветовому тону $\Delta\lambda$ определяет минимальное приращение длины волны, при котором глаз начинает ощущать разницу цветового тона.

Чувствительность глаза к различению цветового тона зависит от области спектра: минимальные значения порога $\Delta\lambda = 1...2$ нм на интервале длин волн видимого диапазона 0.43 ...0.52 мкм. Хуже различаются цветовые тона на границах видимого диапазона, где $\Delta\lambda$ достигает 4...6 нм и более.

Если принять в качестве среднего значения порог по цветовому тону $\Delta\lambda = 2$ нм, то в видимой области спектра от 700 нм до 400 нм (700-400=300 нм) получим число n_λ различимых глазом цветовых диапазонов. При постоянной яркости $n_\lambda = 300: 2 = 150$.

Чистота цвета (уровень цветового тона) устанавливает относительное содержание яркости монохроматического излучения B_λ на данной длине волны к яркости белого излучения B_σ :

$$\rho = \frac{B_\lambda}{B_\lambda + B_\sigma}. \quad (3)$$

Если добавка белого излучения соответствует ($B_\sigma=0$), то монохроматическое излучение на волне λ имеет максимальную насыщенность. Изменение насыщенности цвета обнаруживается глазом при некотором минимальном различии чистоты $\Delta\rho$ - порога цветоразличения по насыщенности:

$$\Delta\rho = \rho_1 - \rho_2, \quad (4)$$

где

ρ_1, ρ_2 – сравниваемые значения чистоты цвета.

Наибольшая чувствительность глаза к изменению чистоты цвета находится в желтой области спектра ($\lambda = 550...580$ нм). Здесь число порогов n_ρ , необходимое для цветоразличения по чистоте, равно шести. На краях видимого диапазона оно достигает 18.

В среднем по спектру $n_\rho = 15$.

Число градаций яркости (светлоты) n_B в диапазоне от самого темного до наиболее светлого цвета при данной освещенности определяется из зависимости от порогового контраста:

$$n_B = 2 + \frac{\lg \varepsilon}{\lg(1 - \varepsilon)}, \quad (5)$$

где

ε - пороговый яркостный контраст.

При $\varepsilon = 0.035$ (пороговый контраст при наблюдении объектов на местности) число градаций яркости n_B равно 96.

В результате получаем, что общее число цветов, различимых глазом:

$$M = n_\lambda \times n_\rho \times n_B = 150 \times 15 \times 96 = 216000. \quad (6)$$

Таким образом, количественная (классическая) теория информации, разработанная для поддержки технических систем связи прошлого века, положила начало формированию новых методов математической статистики и теории вероятности применительно к ее приложениям в различных областях науки и современным системам, как техническим, так и организационным. Необходимым условием использования классических основ теории информации в конкретном приложении, остается уточнение термина «информация» применительно к данной области исследований.

С развитием технических систем и глобализацией коммуникационных связей возникает острая необходимость в разработке новых алгоритмов количественной оценки информации, предполагающих учет не только статистических связей элементов сообщения, но и разнопараметрические оценки этих связей с учетом системного анализа.

Возникающие в различных областях науки «забавные парадоксы теории информации» /2/, являются свидетельством не ограничений классической теории информации (называемой иногда «математической теорией связи»), а следствием прямого перенесения положений теории информации и применения математического аппарата количественной оценки без глубокого системного анализа информационных процессов.

Список литературы

1. Шеннон К. Работы по теории информации и кибернетике (перевод с английского, под редакцией Р.Л. Добрушина и О.В. Лупанова). – М.:Изд. иностр. лит., 963. – С.667-668).
2. Э.А. Соснин. Классическая теория информации и ее ограничения. 2002 г.
3. Корогодин В.И., Соснин Э.А., Пойзнер Б.Н. Рабочая книга по социальному конструированию (Междисциплинарный проект). Ч. 1. Томск: Изд-во Том. ун-та, 2000. 152 с.
4. Луизов А.В., Федорова Н.С. Глаз как приемник информации. В кн. «Специальные вопросы светотехники в охране труда». М.: Наука, 1975. - 250 с.