

Концептуальная оценка качества информации, циркулирующей в организационно-технических системах

В статье рассматривается подход к оценке качества информации, получаемой в процессе преобразования исходной информации, полученной по сигналам управления в организационно-технических системах, в управляющие решения, составляющие содержание процесса управления. Сущность подхода заключается в том, качество информации предлагается оценивать по ее количественной характеристике (относительного количества синтаксического, семантического и прагматического видов информации, находящихся во взаимосвязях и взаимоотношениях).

Ключевые слова: теория информации, виды информации, методы оценки качества информации.

THE CONCEPTUAL QUALITY EVALUATION OF THE INFORMATION CIRCULATING IN TECHNICAL-ORGANIZATIONAL SYSTEMS

The article discusses the approach to assessing the quality of the information obtained in the process of transformation of the initial information received by control signals in the organizational and technical systems, control solutions, make up the content management process. The essence of the approach is to assess the quality of the information offered on its quantitative characteristics (the relative amount of syntactic, semantic and pragmatic information types are in relationships and relationships).

Keywords: information theory, types of information, methods for assessing the quality of information.

1. Введение

Управление в организационных и технических системах и информация тесно связаны между собой. Управление представляет собой процесс упорядочения управляемых объектов, а информирование в организационно-технической системе это процесс упорядочения отражения знаний. Действительно, центральное место в процессе управления (на любом уровне) занимает событие выбора лицом, принимающим решение (ЛПР) действия из некоторого множества возможных вариантов действий с их специфическими исходами. Выбор ограничивает неопределенность поведения исполнителя (активного элемента системы управления), так как он обусловлен появлением информации. Вырабатывая информацию, система управления противостоит тенденции к возрастанию энтропии. Поскольку акт управления является целе-

направленным (например, на увеличение коэффициента готовности технической системы), то он является актом принятия решения.

Известно, что постоянное увеличение объемов передаваемой информации в организационных системах не приводит к такому же росту качества управления, а зачастую приводит к его снижению. Это объясняется отсутствием необходимого условия – мотивации для восприятия (усвоения сущности) информации приемником (исполнителем), а также достаточного условия – сходства (в пределе, совпадения) целей источника информации (руководителя) и приемника (исполнителя). Таким образом, возникает проблема единства понимания сигналов управления источником и потребителем на уровне символов.

Преобразование исходной информации, полученной по сигналам управления, в управляющие решения и составляет содержание

процесса управления. При этом очень важным является решение задачи оценки качества информации, циркулирующей в организационно-технических системах.

Различают четыре основных составляющих (аспекта) информации: синтаксический (статистический, энтропийный), семантический (тезаурусный), прагматический (статистический, целевой) и машинный (структурный).

Последний аспект – структурный связан с преобразованием информации в машинную форму, то есть с проблемами хранения, реструктуризации и извлечения информации, что является чрезвычайно важным при переходе от банков данных к базам знаний, так как он позволяет на логическом уровне определять структуру информационной базы. В то же время при структурном анализе преобразования информации не учитываются аналитические и прагматические аспекты, так как



Геннадий Александрович Беркетов,
к.т.н., профессор
Российский экономический
университет им. Г.В. Плеханова
Тел.: 8 (495) 442-61-11
Эл. почта: GABerketov@mesi.ru

Gennady A. Berketov,
PhD in Technical Sciences, Professor
Plekhanov Russian University
Tel.: 8 (495) 442-61-11
E-mail: GABerketov@mesi.ru



Андрей Александрович Микрюков,
к.т.н., доцент
Российский экономический
университет им. Г.В. Плеханова
Тел.: 8 (495) 442-61-11
Эл. почта: AMikrukov@mesi.ru

Andrey A. Mikrukov,
PhD in Technical Sciences,
Associate Professor
Plekhanov Russian University
Tel.: 8 (495) 442-61-11,
E-mail: AMikrukov@mesi.ru

обработка информации происходит по «машинным законам», вне зависимости от ее смыслового содержания. Поэтому, решение данной задачи выходит за пределы рассматриваемой проблемы.

Для оценки качества получаемой информации предлагается оценивать количественную характеристику качества информации по содержанию составляющей, необходимой для превращения ее в целевое действие, то есть предлагается оценивать относительные количества синтаксического, семантического и прагматического видов информации, находящихся во взаимосвязях и взаимоотношениях.

2. Решение задачи

Появление синтаксической информации связано с преобразованием (обработкой) данных, когда они превращаются в информацию. Основными видами обработки данных при этом являются помехоустойчивое кодирование, шифрация и внесение других видов информационной избыточности, в том числе, позволяющей максимально использовать тезаурус потребителя информации, а значит повысить количество семантической информации, содержащейся в синтаксической информации.

При рассмотрении синтаксического вида информации учитывается не столько сущность собственно информации, сколько ее материального носителя. Это объясняет вероятностную трактовку синтаксической информации. При рассмотрении синтаксической информации объектом исследования является ее структура и виды сигналов ее представления. При этом под сигналом понимается сообщение о любом

изменении начального состояния объекта, которое может вызвать реакцию человека или прибора. Слово «может» подчеркивает, что эту реакцию вызывает не вся синтаксическая информация, а ее прагматическая часть.

Энтропийный подход, являющийся важной составной частью теории информации, которая занимается изучением количества информации в сообщениях безотносительно их конкретного содержания, так как процесс формализации и автоматизации обработки информации (передачи, хранения, переработки) не предусматривает изменение содержания сообщений – семантики.

Поэтому, учет синтаксической информации, используемый в теории передачи информации, является необходимым, но недостаточным условием для корректной оценки качества передаваемой информации в форме знаний, а затем и превращении их в действие.

При построении моделей информационных отношений принято выделять три уровня: концептуальный, логический и физический.

Физический уровень связан со способом реализации информационной модели в техническом устройстве и не имеет отношения к рассматриваемой задаче.

На логическом уровне (формализация информационной модели) решается задача определения аналитического описания каждой из составляющих информации и их отношений.

На концептуальном уровне (описание содержания и структуры предметной области) достаточно рассмотреть бинарные отношения между синтаксической, прагматической и семантической информациями, которые являются частично упорядоченными парами. Иллюстрацию

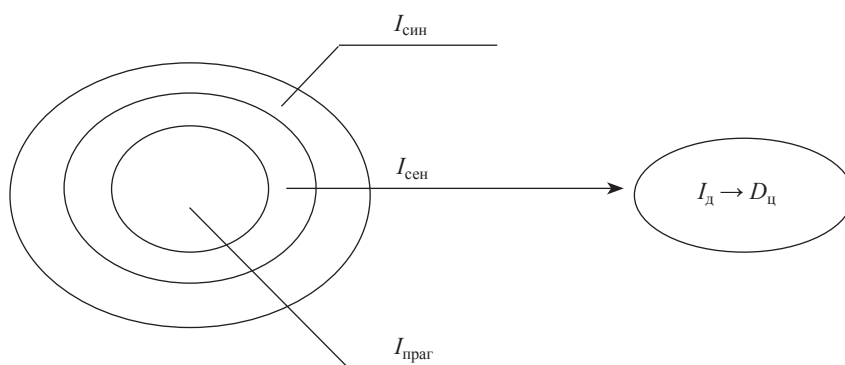


Рис. 1. Отношения доминирования основных составляющих информации



Анатолий Иванович Полоус,
к.т.н., доцент
Российский экономический
университет им. Г.В. Плеханова
Тел.: 8 (495) 442-61-11
Эл. почта: APolous@mesi.ru

Anatolij I. Polous,
PhD in Technical Sciences,
Associate Professor
Plekhanov Russian University
Tel.: 8 (495) 442-61-11,
E-mail: APolous@mesi.ru

отношений доминирования разных составляющих информации можно представить графически на плоскости (рис. 1), но плоское представление не является конструктивным, так как не позволяет задавать качество информации. Пространственное представление (рис. 2) позволяет решить проблему, по меньшей мере, на концептуальном уровне.

В терминах теории множеств информацией об образе в разных аспектах информации ($I_1 = I_{\text{син}}$, $I_2 = I_{\text{сем}}$, $I_3 = I_{\text{праг}}$, $I_4 = I_{\text{дейс}}$) являются разные совокупности характеристик объекта, множества которых не являются собственными подмножествами друг для друга.

$$I_{\text{син}} \supseteq I_{\text{сем}} \supseteq I_{\text{праг}} \supseteq (I_{\text{д}} \leftrightarrow D_{\text{ц}}) \quad (1)$$

В выражении (1) целевая составляющая информации $I_{\text{д}}$, воплощаемая в действие $D_{\text{ц}}$ (целое или часть прагматической информации $I_{\text{праг}}$), является энергетической составляющей информации, так как она непосредственно воздействует на объект, и является той ее частью, которая, по словам Вернадского, «начинает воздействовать на костную материю».

Из рис. 1 следует, что повышение качества информации соответствует максимальному сближению (совпадению) трех окружностей.

Трехмерное пространство бинарных отношений между синтаксической, прагматической и семантической информацией в форме прямоугольного параллелепипеда (рис. 2) позволяет ввести понятие качества информации на концептуальном уровне. Такое совместное представление аспектов информации в виде «информационного параллелепипеда», образованного тремя координатами $I_{\text{син}}$, $I_{\text{сем}}$, $I_{\text{праг}}$ становится достаточно конструктивным, так как качество информации можно оценивать мерой близости информационного параллелепипеда к кубу по форме при условии со-

хранения равенства объемов этих фигур $V_{\text{пар}} = V_{\text{куб}}$. В такой трактовке качество информации будет повышаться по мере приближения прямоугольного параллелепипеда к кубу по форме (рис. 2) при условии $V_{\text{пар}} = V_{\text{куб}}$, где $V_{\text{пар}} = I_{\text{син}} \cdot I_{\text{сем}} \cdot I_{\text{праг}}$; $V_{\text{куб}} = I_{\text{син}}^3 = I_{\text{сем}}^3 = I_{\text{праг}}^3$.

Из рис. 2 следует, что, чем ближе форма информационного параллелепипеда с ребрами $I_{\text{син}} \leq I_{\text{праг}} \leq I_{\text{сем}}$ (при $V_{\text{прям}} = V_{\text{куб}}$) к кубу, ($I_{\text{син}} = I_{\text{праг}} = I_{\text{сем}}$), тем выше качество $K(I)$ информации в том смысле, что в синтаксической информации содержится максимально возможная часть прагматической информации, непосредственно используемой потребителем для совершения целевого действия.

Таким образом, качество информации максимально, $K(I) \rightarrow K(I)_{\text{max}}$, при условии $(\Delta_{\Sigma} = \Delta_1 + \Delta_2 + \Delta_3) \rightarrow 0$, где $\Delta_1 = I_1 - I_2$; $\Delta_2 = I_2 - I_3$; $\Delta_3 = I_3 - I_{\text{д}}$, так как в этом случае семантическая (смысловая) составляющая занимает большую часть синтаксической информации (в пределе равна ей), а прагматическая (целевая) составляющая – семантической информации (в пределе равна ей).

Тогда, критерием качества $K(I)$ информации на концептуальном уровне может служить своеобразный «коэффициент полезного действия информации», задаваемый минимизацией суммы Δ_{Σ} попарных разностей Δ_1 , Δ_2 , Δ_3 , путем варьирования количественных значений составляющих информации $I_{\text{син}}$, $I_{\text{сем}}$, $I_{\text{праг}}$

$$K_{\text{max}}(I) \rightarrow \text{Min} \Delta = \text{Min}(\Delta_1 + \Delta_2 + \Delta_3), \quad (2)$$

$I_1 \leq I_2 \leq I_3$ $I_1 \leq I_2 \leq I_3$

где: $\Delta_1 = \Delta_{\text{син}} = I_{\text{син}} - I_{\text{сем}} = I_1 - I_2$;
 $\Delta_2 = \Delta_{\text{сем}} = I_{\text{сем}} - I_{\text{праг}} = I_2 - I_3$;
 $\Delta_3 = \Delta_{\text{праг}} = I_{\text{праг}} - I_{\text{дейс}} = I_3 - I_{\text{д}}$.

Все слагаемые в (2) в соответствии с физической интерпретацией не равны нулю и положительны. Как известно, количество инфор-

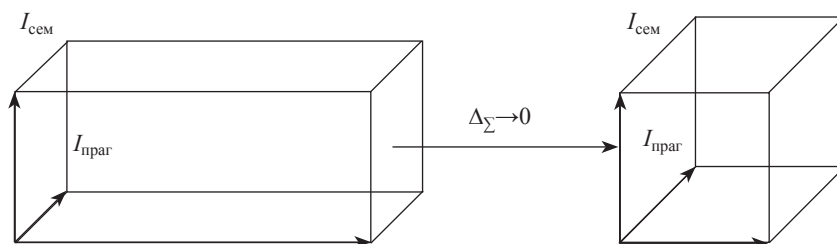


Рис.2. Пространственное представления отношений доминирования основных составляющих информации

мации может быть нулевым только при бессмысленных сообщениях (типа белого шума) [1]. Из выражения (2) видно, что чем больше в структурной части ($I_{\text{син}}$) информации, передаваемой от источника A к потребителю B , знаний ($I_{\text{сем}}$), а в знаниях – прагматической ($I_{\text{праг}}$) (целевой) информации, тем выше качество информации, так как для выполнения целевого действия ($I_{\text{д}}$) будет использоваться максимально возможная часть всей переданной структурной информации $I_{\text{син}}$.

В семантической теории информации используются ряд вероятностных подходов к определению качества синтаксической информации (количества семантической информации) [1].

Качество информации, как меру ее целесообразности можно оценивать разницей вероятностей достижения цели после получения информации p_1 и до ее получения p_0 [4]:

$$K(I) = \log_2 p_1 - \log_2 p_0 = \log_2 p_1 / p_0.$$

Известны более поздние подходы [1] к оценке качества информации через содержательность $\text{cont}(i)$ события i , выражаемую логическими функциями истинности $m(i)$ и ложности $m(i)$, которые сходны с вероятностями событий $p(i)$ и анти-событий $p(i)$, ($p(i) + m(i) = 1$).

$$K(I) = \text{cont}(i) = m(i) = 1 - m(i).$$

В некоторых работах качество $K(I)$ информации вводится через понятия статистического $I_{\text{стат}} = \log_2 1/p(i) = -\log_2 p(i)$ и логического $I_{\text{лог}} = \log_2(1/(1 - \text{cont}(i))) = \log_2 1/m(i) = -\log_2 m(i)$ количества информации $K(I) = K(I_{\text{стат}}, I_{\text{лог}})$. При этом статистическая оценка учитывает вероятность реализации события i , а логическая – меру его истинности или ложности.

Для случаев, когда целью извлечения информации является обращение неопределенности ситуации в ноль, например, при диагностировании, ее можно представить

множеством отношений предшествования между причиной (а) и следствием (б) ($a_i \rightarrow b_j$) с вероятностью p_{ij} . Для множества следствий $N(t)$ и причин $M(t)$ в момент t энтропия определяется выражением

$$H(t) = - \sum_{i=1}^{N(t)} \sum_{j=1}^{M(t)} p_{ij}(t) \log p_{ij}(t),$$

а для момента $(t + 1)$ выражением:

$$H(t+1) = - \sum_{i=1}^{N(t+1)} \sum_{j=1}^{M(t+1)} p_{ij}(t+1) \log p_{ij}(t+1).$$

Тогда качество информации на интервале $[t, t + 1]$ можно выразить через динамическую энтропию $K(I) = H(t) - H(t + 1)$, которое может быть нулевым или отрицательным (если неопределенность в момент $(t + 1)$ не уменьшилась).

Это – вероятностные (энтропийные) оценки, которые не связаны с содержанием знаний в передаваемой информации и не рассматривают обработку знаний.

Учитывая тот факт, что логическому мышлению, логическим выводам, логическим решениям, нормальной психике, интеллекту, интуиции, отсутствию информации свойственна нулевая энтропия [2], для оценки качества семантической информации воспользуемся тезаурусным подходом.

Логический уровень. Для формального представления предложенной информационной модели (на логическом уровне) необходимо определить аналитические выражения для частных оценок качества $K(I_{\text{син}})$, $K(I_{\text{сем}})$, $K(I_{\text{праг}})$ каждой из составляющих информации. Это необходимо для гармонизации информационного ресурса системы передачи информации. Известно [3], что алгоритм передачи информации при управлении сложными организационными системами должен строиться исходя из постулата гармонизации канала передачи знаний, обеспечивающего наиболее рациональный расход количества информационного ресурса. При

этом необходима привязка получаемых значений к конкретному моменту времени, так как ценность информации со временем падает.

В соответствии с вышеизложенным, качество $K(I)$ информации предлагается оценивать по формуле

$$K(I) = 1 - \Delta_{\Sigma} / \Delta_{\Sigma_{\text{max}}}, \quad (3)$$

где $\Delta_{\Sigma_{\text{max}}} = I_{\text{син}} - I_{\text{д}}$;
 $K(I) \rightarrow 1$ при $\Delta_{\Sigma_{\text{max}}} \rightarrow 0$;
 $K(I) \rightarrow 0$ при $\Delta_{\Sigma} = \Delta_{\Sigma_{\text{max}}}$.

Заключение

Известные методы оценки качества информации как меры ее целесообразности используют вероятностный подход (разницу вероятностей достижения цели после получения информации p_1 и до ее получения p_0), либо статистический подход (вероятность реализации события или логическая мера его истинности или ложности), которые слабо отражают связь с факторами, определяющими качество информации и являющимися в определенной мере артефактами.

Наиболее понятным и естественным является предлагаемый подход к определению качества информации, циркулирующей в организационных и технических системах, учитывающий легко определяемое количество каждого вида информации и их соотношения.

Обычно при передаче информации на физическом уровне, из соображений скрытности, помехоустойчивости и других требований к каналу, количество $I_{\text{сем}}$, содержащееся в $I_{\text{син}}$, преднамеренно занижается в связи с использованием в передаваемой кодовой комбинации служебной информации (например, при помехоустойчивом кодировании). Это – один из методов использования информационной избыточности канала, следствием чего является запланированное снижение качества $I_{\text{син}}$, которое легко может быть учтено при организации информационного канала.

Литература

1. Теоретические основы информатики и информационная безопасность. // Под ред. В.А. Минаева, В.Н. Саблина. – М.: Радио и связь, 2000.
2. Прангшвили И.В. Энтропийные и другие системные закономерности Вопросы управления сложными системами. – М.: Наука, 2003.
3. Хайкин Саймон. Нейронные сети: полный курс, 2-е издание: Пер. с англ. – М.: Изд.дом «Вильямс», 2006.
4. Харкевич А.А. О ценности информации. //Проблемы кибернетики, вып. 4, 1961.