

УДК 004.77

ТЕОРЕМА БАЙЕСА В КОГНИТИВНОЙ СЕМАНТИКЕ ОБРАЗОВАТЕЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Болбаков Р. Г.

Московский государственный технический университет радиотехники, электроники и автоматики (МГТУ МИРЭА)

В работе рассматривается раскрытие таких понятий как «когнитивность» и «семантика», а также их соединение в направлении информационных образовательных системах. В статье впервые делается попытка объединить смысловую оценку сетевого отклика, его семантику с оценкой самого акта познания и приращения знаний (негэнтропии знания) в одном понятии – когнитивная семантика. Формализацию этого нового понятия автор предлагает выполнить с привлечением вероятностной теории Байеса (Bayes). Предметом исследования когнитивной семантики образовательной информации автором выбрана модель информационного поля портално-сетевого консорциума, находящегося в распоряжении объекта обучения. На основе теории Байеса производится моделирование когнитивной семантики сетевого микропортального консорциума. В качестве примера приводится решение задачи определения когнитивности ресурса трех вертикальных образовательных порталов. Основные существенные результаты заключаются в получении детерминированного метода прогнозирования когнитивной образовательной информации в консорциуме на основе теоремы Байеса как регулятора когнитивной семантики в информационном образовательном поле для сбора и ведения статистики когнитивной семантики.

Ключевые слова: семантика, пертинентность, релевантность, когнитивность, когнитивная семантика.

BAYES THEORY IN COGNITIVE SEMANTICS OF EDUCATIONAL INFORMATION SYSTEMS

Bolbakov R. G.

Moscow State Technical University of Radio Electronics and Automation

In the introduction to the paper a good outlook for the cognitive semantics in educational information technologies is noted. Modeling of cognitive semantics for a portal consortium is performed on the basis of Bayes theorem. As an example, the solution to the problem of resource cognitivity estimation for three vertical educational micro portals is presented. The problem of determination of cognitive files has been considered on the basis of generalized entropy theory. If an event "A" come at the presence of only one of incompatible events (hypothesizes), the total probability formula can be presented in form of the graph with an isolated vertex "A". The conditional probabilities for the events (hypothesizes) are calculated according to the Bayes formulae. It is noted that the Bayes theory provides a way of revising the hypothesis probability with regard to experimental results or concrete practice. Also possible solutions for two additional problems are presented: 1) estimation of probability of any randomly selected file cognitivity, and 2) which net portal contains a randomly selected file turned out to be noncognitive. The procedures presented in the paper may be useful at practical determination of the cognitivity of educational resources in both standard and remote communications.

Key words: semantic, pertinence, relevance, cognitivity, cognitiv semantics.

В теории информационных процессов и систем термин «когнитивность» используется в широком смысле, обозначая меру достижения познания или, что тоже самое, приращение знания (негэнтропию знания) [8]. В такой постановке этот термин может быть интерпретирован в дидактическом плане как обозначающий появление и «становление» знания и концепций, связанных с этим знанием, выражающих себя как в мысли, состоянии, так и в действии. Нередко термин «когнитивность» употребляется в процессах изучения так называемого «контекстного знания», т.е. в предметных областях, где фигурируют такие понятия как знание, умение или обучение. Извлечение и применение контекстных знаний опираются на понятие семантики, являющей формализацию ключей к этим знаниям. Именно

с этих позиций, а также с учётом воздействия технологических факторов автор настоящей статьи оценивает возможности моделирования и регулирования когнитивной семантики информационных процессов и систем, обслуживающих образовательную индустрию. Цель настоящей работы – показать на конкретном примере продуктивность этой парадигмы и дать её развитие до уровня наглядного модельного описания и рекомендаций.

Под термином «семантика» понимают [2, 5, 9] интерпретацию связи содержания с формой, систему правил определения поведения отдельных языковых конструкций. Объединение понятий «когнитивность» и «семантика» позволяет не только формализовать субъективную сторону познания, но также выработать оценку семантической конструкции, в том числе по принципу полноты и «удобства» её восприятия конкретным субъектом. Отсюда возникает обобщённое обновляющее понятие «когнитивная семантика». Когнитивная семантика может служить достаточно перспективным, ёмким и унифицированным мерилем и инструментом регулирования функциональных достоинств информационных систем в образовании, в том числе электронных библиотек, систем дистанционного обучения и консорциумов образовательных порталов. Плодотворность этой парадигмы подтверждена на примерах многоуровневых порталных консорциумов, исследованием, созданием и сопровождением которых автор длительное время занимался в рамках соответствующих отраслевых высшей школы и московских научно-производственных образовательных программ.

В предшествующей публикации автора [1] и отчётах по соответствующим НИР и НИОКР, проведенных в МГТУ МИРЭА в период 2008–2011 гг. показано, что практический интерес в образовательных информационных технологиях представляет применение методов когнитивной семантики, для количественной оценки вероятности извлечения «когнитивного документа» из совокупного информационного ресурса сетевого порталного консорциума, а также для определения вероятности, с какой этот документ может находиться в одной из подсистем этого консорциума. Здесь термин «когнитивная семантика» трактуется и применяется в уточнённом дидактическом смысле, то есть как оценка истинности и комфортности обнаружения поиска, извлечения, полноценного раскрытия и восприятия объектом обучения откликов информационной системы на его запросы, причём откликов, содержащих истинную знаниевую информацию.

В семантических системах пертинентность ответов ИС (точное и полное соответствие содержания выданного документа сути запроса) автоматически означает его релевантность (соответствие содержания документа формулировке запроса по формальным признакам, в частности, идентификаторам мультимедиа контейнера, файла или атома в составе HTML файла). Однако не все выданные системой документы, имеющие признаки пертинентности, с

позиции восприятия их конечным пользователем «дружественны», доступны быстрой оценке и реально пригодны к использованию. Иными словами нарушаются мажоритарный и эргодический принципы функционирования дидактических информационных систем, условно отнесённые выше к термину «комфортность» работы со знаниевой информацией в составе информационных образовательных консорциумов. Из совокупности пертинентных ответов системы пользователь по его собственным критериям отбирает наиболее полезную и удобную для себя информацию как по форме, так и по содержанию, то есть наиболее мажоритарную. Таким образом, когнитивную семантику конкретного документа, другими словами, его когнитивность, в реальной образовательной практике представляется возможным определить не только по соответствию выданного документа формализованному на уровне соблюдения релевантности запросу, но также с учётом компетентностных и психо-эмоциональных особенностей пользователя. Формализацию этого нового понятия предлагается выполнить с привлечением хорошо развитого математического аппарата описания обширного разнообразия энтропийных оценок информационного морфизма, т.е. функционала взаимодействия, информационных систем между собой и с конечным пользователем [1]. При этом опора делается на широко известный авторитетный подход на основе вероятностной теории Байеса – Колмогорова [6–8, 10].

В качестве типового примера рассмотрим задачу моделирования когнитивной семантики, базируясь на априорном знании когнитивности образовательной информации в информационном ресурсе портално-сетевом консорциума, включающего на аддитивной основе три образовательных микропортала [1]. Полный ресурс консорциума в эксперименте составляет 10000 элементарных семантических единиц (ЭСЕ), отобранных случайным образом из образовательного портала, включая до 10 % системной и сопутствующей информации, которая не учитывается в решении задачи. При этом первый из микропорталов содержит 25 % совокупного информационного ресурса, второй – 35 %, третий – 40 %. При этом методом экспертных оценок в практических упражнениях преподавателем определена потенциальная когнитивность ЭСЭ для каждого из микропорталов, соответственно 5, 4 и 2 %. (заметим, что в многократных опытах подобного рода образующиеся численные значения были схожими с приведёнными выше). Доступ к образовательным ресурсам микропорталов осуществляется последовательным запросом каждого из ЭСЭ без повторения. Отметим также, что сбор статистики запросов к микропорталам или другим образовательным ИС в реальных образовательных условиях осуществляется за определенный отрезок времени, будь то день, неделя, месяц или год, абстрагируясь от чего в ходе моделирования, запросы к микропорталам задаются последовательно и не зависят от времени. Сведения о ресурсах, участвующих в эксперименте микропорталов, сведены в таблицу 1.

Таблица 1. Распределение ЭСЕ в микропорталах сетевого консорциума

Микропорталы	ЭСЕ	Кол-во ЭСЕ микропорталов	Когнитивность /ЭСЕ
1	25 %	2500	5 % / 12,5
2	35 %	3500	4 % / 14
3	40 %	4000	2 % / 8

Далее, задачу определения вероятности получения отклика ИС в виде когнитивного файла (ЭСЕ) будем решать по формуле полной вероятности [1, 2].

Если событие A наступает лишь при появлении одного из несовместных событий (далее гипотез) H_i , то вероятность события A вычисляется по формуле полной вероятности:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + \dots + P(H_n)P(A/H_n), \quad (2)$$

где $P(H_i)$ – вероятность гипотезы H_i ,

A – событие, которое появляется после появления событий H_i . В нашем случае A – получение дефектного (некогнитивного) файла из числа образовательных ЭСЕ ($P(A/H_i)$), с выявлением вероятности поддержания одним из трех микропорталов ($P(H_i)$) порталного-сетевого консорциума;

$\sum_{i=1}^n P(H_i) = 1$, – условная вероятность события при выполнении гипотезы H_i ($i = 1, 2, \dots, n$).

Представим формулу полной вероятности графом с выделенной вершиной A :

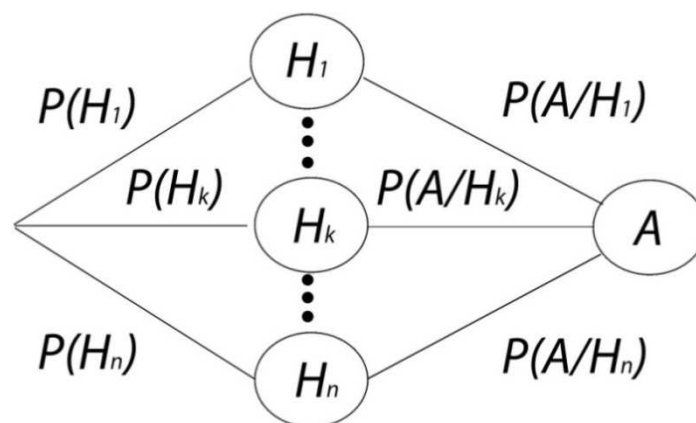


Рис. 1. Граф полной вероятности

Полная вероятность события A равна весу всего вероятностного графа с гипотезами.

Теперь рассмотрим ресурс образовательного консорциума с позиции теории Байеса [8].

Если до опыта вероятности гипотез были $P(H_1)$, $P(H_2)$, ..., $P(H_n)$, а в результате опыта появилось событие A , то с учетом этого события «новые», т.е. условные вероятности гипотез вычисляются по:

$$P(H_k / A) = \frac{P(H_k)P(A / H_k)}{P(A)}, \quad (3)$$

$$\text{где } (k = 1, 2, \dots, n) \text{ и } P(A) = \sum_{i=1}^n P(H_i)P(A / H_i). \quad (4)$$

Включая описываемый здесь фрагмент в регулирование когнитивности образовательных порталов, автор исходит из того, что в общепринятом представлении теорема Байеса (формула Байеса) [7, 8] дает возможность «пересмотреть» вероятность гипотез с учетом наблюдавшегося результата опыта. Условная вероятность $P(H_k / A)$ может находиться как отношение веса ветви, проходящей через вершину, соответствующую гипотезе H_k , к весу всего вероятностного графа.

В объявленную выше постановку определения вероятности получения отклика ИС в виде когнитивного файла (ЭСЕ) могут быть добавлены две дополнительные развивающие задачи:

- Оценить вероятность того, что случайно вызванная ЭСЕ когнитивна.
- Определить, какому микропорталу принадлежит случайно выбранный из консорциума элемент, оказавшийся некогнитивным.

Ниже приводятся следующее возможное решение этих задач.

Пусть рассматривается событие $A = \{\text{выбрать дефектный файл}\}$. Сформулируем три возникающие в процессе рассмотрения гипотезы:

$$H_1 = \{\text{файл поддерживается первым микропорталом}\}, P(H_1) = 0,25 \text{ (25 \% - 2500 ЭСЕ)},$$

$$P(A / H_1) = 0,05 \text{ (5 \% ЭСЕ первого микропортала)};$$

$$H_2 = \{\text{файл поддерживается вторым микропорталом}\}, P(H_2) = 0,35 \text{ (35\% - 3500 ЭСЕ)},$$

$$P(A / H_2) = 0,04 \text{ (4 \% ЭСЕ второго микропортала)};$$

$$H_3 = \{\text{файл поддерживается третьим микропорталом}\}, P(H_3) = 0,4 \text{ (40\% - 4000 ЭСЕ)},$$

$$P(A / H_3) = 0,02 \text{ (2 \% ЭСЕ третьего микропортала)}.$$

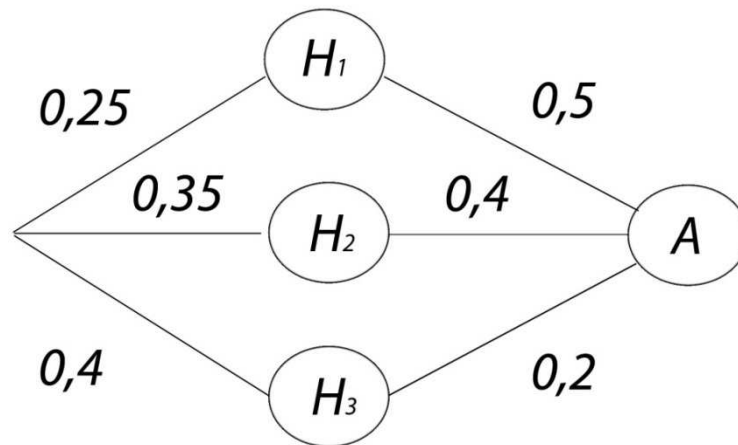


Рис. 2. Граф выдвигаемых гипотез

$$P(A) = 0,25 \cdot 0,05 + 0,35 \cdot 0,04 + 0,4 \cdot 0,02 = 0,0345 \quad (5)$$

Если полная вероятность равна $P(A) = 1$, то каждая случайно вызванная ЭСЕ будет когнитивна для всех 9000 образовательных ЭСЕ консорциума. В нашем случае $P(A) = 0,0345$, т.е. с вероятностью 0,0345 ЭСЕ консорциума окажется когнитивным, тогда примерно 34 ЭСЕ из всего ресурса образовательного портально-сетевого консорциума могут оказаться когнитивными.

$$P(H_1/A) = \frac{0,25 \cdot 0,05}{0,0345} = \frac{25}{69} \approx 0,36$$

$$P(H_2/A) = \frac{0,35 \cdot 0,04}{0,0345} = \frac{28}{69} \approx 0,41$$

$$P(H_3/A) = \frac{0,4 \cdot 0,02}{0,0345} = \frac{16}{69} \approx 0,23$$

(6)

Результаты расчетов (5) и (6) показывают, что примерно каждая 345 ЭСЕ может быть когнитивна при $P(A) = 0,0345$ и, с большей вероятностью, будет принадлежать второму микропорталу, равному 0,41, поскольку $P(H_2/A) > P(H_1/A) > P(H_3/A)$. Приведенная процедура расчета вероятности когнитивности извлекаемого ЭСЕ может оказаться применимой в расчетах вероятности получения pertinentных откликов конкретной образовательной ИС. Процедура успешно и без единого сбоя применялась автором в течение четырёх лет в процессе создания и сопровождения упоминавшихся выше многоуровневых образовательных консорциумов. Данная процедура может быть полезна при практическом определении когнитивности ресурса образовательных информационных систем в обычных и дистанционных формах обучения.

Список литературы

1. Болбаков Р. Г. Математическое описание когнитив-энтропии макромедиа образовательных систем. Динамика неоднородных систем / Под ред. Ю. С. Попкова. Вып. 14. – М.: ЛЕНАНД, 2010. – С. 252–260. (Труды Института системного анализа РАН; Т. 53 (2)). Регистрационный номер ПИ № 77-14194. ISBN 978-5-397-00676-7.
2. Казаков И. А., Манцивода А. В. Базы данных как онтологии // Известия Иркутского государственного университета. Серия: Математика. – 2011. – Т. 4. № 1. – С. 20–30.
3. Куракин Д. В., Шемончук Д. С. Расширение использования мультимедиа технологий в сфере образования // Информатизация образования и науки. – 2009. – № 4. – С. 43–50.
4. Лийв Э. Х. Инфодинамика. Обобщённая энтропия и негэнтропия. – Таллинн, 1998. – 200 с. Библ. 131 ед.
5. Малых А. А., Манцивода А. В. Онтологии, метаданные и семантическое программирование // Вестник Новосибирского государственного университета. Серия: Математика, механика, информатика. – 2007. – Т. 7. – № 2. – С. 29–51.
6. Хренников А. Ю. Формула полной вероятности с интерференционным членом и представление колмогоровской модели в гильбертовом пространстве // Теория вероятностей и ее применения. – 2006. – Т. 51. № 3. – С. 518–536.
7. Эддоус М., Стенсфилд Р. Методы принятия решений / Пер. с англ., под ред. член-корр. РАН И.И. Елисеевой. – М.: Аудит, ЮНИТИ, 1997. – С. 22–26.
8. Bayes, Thomas, and Price, Richard, An Essay towards solving a problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F.R.S.] // Philosophical Transactions of the Royal Society of London 53. – 1763. – P. 370–418.
9. Berners-Lee T., Hendler J., Lassila O. The Semantic Web: Overview / Semantic Web // The Scientific American. – 2008. – № 5. – P. 32–45.
10. Kolmogoroff A. N., Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer-Verlag, Berlin, 1933, 62 pp.; Рус. перев.: Колмогоров А. Н. Основные понятия теории вероятностей. – М.: Наука, 1974. – 119 с. MathSciNet.

Рецензенты:

Раев Вячеслав Константинович, доктор технических наук, профессор кафедры ТИССУ ФГБОУ ВПО МГТУ МИРЭА, г. Москва.

Петров Андрей Борисович, доктор технических наук, профессор, декан факультета «Информационные технологии» ФГБОУ ВПО МГТУ МИРЭА, г. Москва.