

УДК 004.056

Е.А. Родин

Математическая модель процесса фильтрации сообщений на основе диффузии материалов в тракте обработки

В автоматизированных системах фильтрации информации существует проблема оценивания качества ее функционирования, связанная со сложностью классификации конечного продукта – информации, а также неопределенностью относительно источников информации. В работе рассматривается математическая модель процесса отбора сообщений на основе диффузии материалов в тракте обработки, учитывающая особенности системы фильтрации – множество классификационных признаков и пропускная способность системы фильтрации.

Ключевые слова: система фильтрации сообщений, диффузия, диффузия материалов в тракте обработки, классификация информации.

Процесс создания и применения любой информационной системы (ИС) неизбежно связан с проблемой оценки качества ее функционирования. Сложность проблемы заключается, во-первых, в свойствах конечного продукта – информации, во-вторых, неопределенностью относительно источников получаемой информации. Показателем эффективности функционирования ИС фильтрации информации является степень достижения цели, т.е. качество классификации (фильтрации) сообщений.

Методы математического моделирования прочно вошли в арсенал познавательных средств, используемых современной наукой. Сущность этой методологии состоит в замене исходного объекта его «образом» – математической моделью и дальнейшем изучении модели с помощью реализуемых на компьютерах вычислительно-логических алгоритмов [1–3]. Работа не с самим объектом (явлением, процессом), а с его моделью дает возможность безболезненно, относительно быстро и без существенных затрат исследовать его основные свойства и поведение в любых ситуациях. В то же время вычислительные (компьютерные, имитационные) эксперименты с моделями объектов позволяют, опираясь на мощь современных вычислительных методов и технических инструментов информатики, подробно и глубоко изучать объекты в достаточной полноте, недоступной теоретическим подходам.

В работе [4] высказано предположение о возможности использования S -образных кривых (логистическая, Гомпертца, модифицированная экспоненциальная и др.) и уравнений типа Лотки–Вольтера для моделирования процессов технологического развития.

Согласно [5] технологический процесс – это часть производственного процесса, содержащая целенаправленные действия по изменению и (или) определению состояния предмета труда. Производственный процесс – совокупность всех действий людей и орудий труда, необходимых на данном предприятии для изготовления и ремонта продукции [6].

Экспериментальные исследования показали, что процесс диффузии, выраженный в виде доли выпуска продукции определенного уровня, или доли фирм, освоивших рынок новой продукции, также описывается логистической кривой или ее модификациями [7–10].

С точки зрения термодинамики диффузия (от лат. *diffusio* – распространение, растекание, рассеивание) – перенос частиц разной природы, обусловленный хаотическим тепловым движением молекул (атомов) в однокомпонентных или многокомпонентных газовых либо конденсированных средах [11]. Согласно закону Фика, масса вещества, переносимого за время dt через элементарную площадку dS , пропорциональна градиенту плотности, площадки dS и времени dt [12]

$$dM = -D \frac{dp}{dx} dS dt, \quad (1)$$

Главной характеристикой диффузии служит плотность диффузионного потока (коэффициент диффузии) – количество вещества, переносимого в единицу времени через единицу площади поверхности, перпендикулярной направлению переноса. Коэффициент диффузии отражает скорость диффузии и определяется свойствами среды и типом диффундирующих частиц.

Современная экономическая теория научно-технического развития и инноваций была создана трудами Й. Шумпетера. Наиболее формализованной ее ветвью является концепция диффузии инноваций, в основе которой лежит уравнение Ферхюльста, описывающее логистический рост и широко используемое при изучении динамики популяций [13].

Логистическая кривая является решением уравнения

$$\frac{dx}{dt} = \lambda \cdot x \cdot (Y - x), \quad (2)$$

где x – зависимый от времени показатель эволюции объекта; λ – параметр, определяющий скорость изменений (тренд эволюции); Y – параметр, определяющий предел изменений в рамках текущего состояния рынка (исследуемой системы).

Диффузия инноваций (diffusion of innovation) – это процесс распространения новшеств в обществе, закономерности распространения новых продуктов, технологий, идей среди потенциальных потребителей (пользователей) с момента их появления [14].

Объектом исследования диффузии инноваций являются: новый или усовершенствованный продукт; новый или усовершенствованный технологический процесс.

Основные параметры (характеристики) диффузии инноваций: коэффициенты инновации; рыночная цена на нововведение; затраты на рекламу; затраты на передачу информации; доход, получаемый от внедрения инновации.

Математический аппарат экономической теории научно-технического развития и инноваций можно с успехом применить для моделирования процесса фильтрации сообщений в тракте обработки системы фильтрации информации:

- в качестве инноваций выступает множество сообщений, представляющих ценность для конечных потребителей информации;
- доход, получаемый от внедрения инновации, определяется как количество правильно классифицированных сообщений в определенный момент времени к общему количеству сообщений, классифицируемых по заданному классификационному признаку;
- коэффициент инновации – коэффициент диффузии в классическом (термодинамическом) понимании;
- количественный и качественный состав признаков классификации в текущий момент времени – затраты на рекламу и передачу информации.

В самом общем виде любая система фильтрации состоит из подсистемы приема, подсистемы классификации и подсистемы распределения отобранной информации, объединенных подсистемой управления. На рис. 1 представлена обобщенная структурная схема системы фильтрации сообщений.

Подсистема приема предназначена для первичной фильтрации информации, получаемой из различных источников, таких как сети Internet, Intranet, СМИ и т.д. Первичная фильтрация может осуществляться на основе классификационных признаков, характеризующих источники информации: название журналов, газет, книг, фирм, сайтов и т.д.

Сообщения, прошедшие через подсистему приема, поступают в подсистему классификации, которая предназначена для конечной фильтрации потока и определения тематической принадлежности сообщений, например: экономика, политика, культура и т.д.

Подсистема распределения предназначена для доставки классифицированных сообщений конечным пользователям. Фактически данная подсистема реализует функции «маршрутизатора» информации.

Подсистема управления предназначена для контроля выполнения технологических процессов и формирования управляющих воздействий в зависимости от требований конечных потребителей информации.

Оказывать влияние на систему фильтрации информации, интересующей конечных пользователей, возможно за счет изменения (модификации) классификационных параметров. Таким образом, модель процесса фильтрации сообщений должна учитывать особенности массива классификационных параметров, которые обеспечивают непосредственный перевод входного потока сообщений во множество сообщений, представляющих интерес для конечных потребителей.

Классификационный (идентификационный) признак [15] – это свойство объекта предметной области, удовлетворяющее определенным требованиям. Каждый объект может быть выделен из множества сходных объектов по совокупности присущих ему свойств. Для этого могут быть ис-

пользованы любые свойства объекта: размер сообщения, наличие вложений и сигнатур, адресные свойства и т.д.

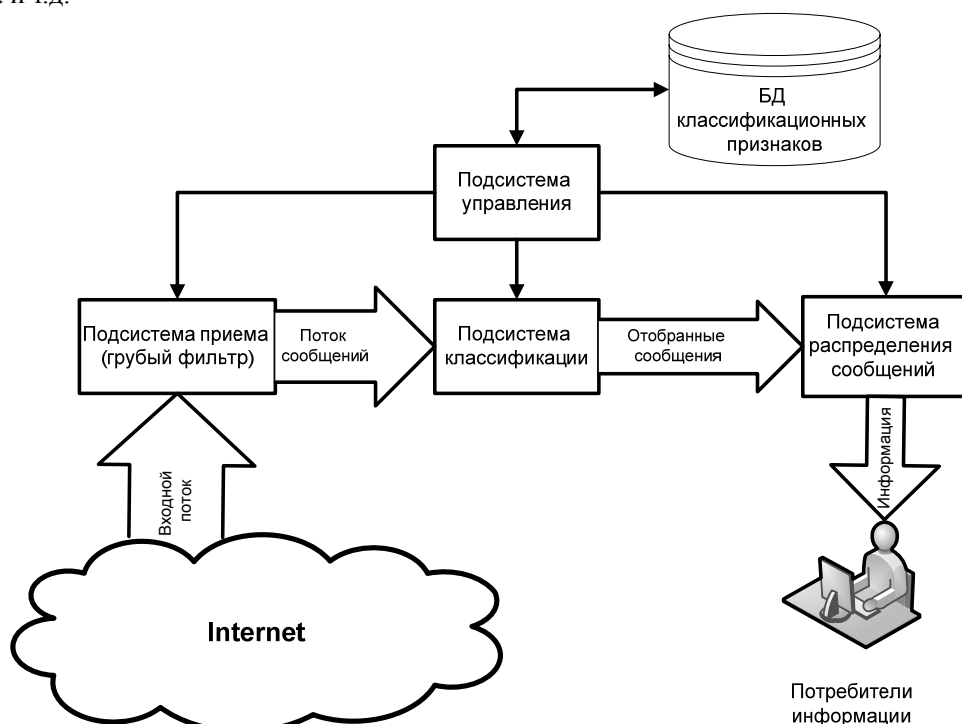


Рис. 1. Обобщенная структурная схема системы фильтрации

Исходное уравнение для модели процесса фильтрации сообщений на основе диффузии материалов в тракте обработки в детерминированном виде имеет вид

$$\frac{dV}{dt} = K(t) \cdot \sum_{i=1}^n W_i(t) \cdot I_i(t), \quad (3)$$

где $i = 1, 2, 3 \dots n$, n – потенциальное количество источников информации; $W_i(t)$ – доля сообщений, представляющих интерес для конечных пользователей информации, i -го источника информации в момент времени t ($1 \geq W_i(t) \geq 0$); $I_i(t)$ – информативность (коэффициент информативности) классификационных признаков i -го источника информации в момент времени t ($1 \geq I_i(t) \geq 0$); $K(t)$ – коэффициент загрузки системы обработки в момент времени t ($1 \geq K(t) \geq 0$).

$$W_i(t) = \frac{x_i(t)}{N(t)}, \quad (4)$$

где $x_i(t)$ – количество сообщений, представляющих интерес для конечных потребителей, i -го источника информации в момент времени t ; $N(t)$ – общее количество сообщений, полученных от подсистемы приема в момент времени t ($Y(t) \geq N(t) > 0$).

$$K(t) = \frac{N(t)}{Y(t)}, \quad (5)$$

где $Y(t)$ – максимально возможное количество сообщений в тракте обработки. Определяется пропускной способностью системы обработки.

Информативность классификационных признаков i -го источника:

$$I_i(t) = \frac{\sum_{j=1}^M \alpha_j(t) \cdot \beta_j(t)}{M_{\text{пр } i}(t)}, \quad (6)$$

где $I_i(t)$ – информативность классификационных признаков i -го источника информации; $M_{\text{пр } i}(t)$ – общее количество признаков классификации i -го источника в момент времени t (мощность призна-

кового пространства i -го источника); $\alpha_j(t)$ – достоверность j -го признака классификации i -го источника; $\beta_j(t)$ – актуальность j -го признака классификации i -го источника.

Все классификационные признаки характеризуются своей актуальностью (коэффициентом актуальности). Коэффициент актуальности – скалярная величина, показывающая степень соответствия признака классификации (фильтрации) текущему моменту времени. Все информационные процессы растянуты во времени и поэтому достоверная и адекватная, но устаревшая информация может приводить к ошибочным решениям [16].

$$\beta(t) = 1 - \frac{p(t)}{p_{\max}}, \quad (7)$$

где $\beta(t)$ – коэффициент актуальности признака классификации на текущий момент времени ($1 \geq \beta(t) \geq 0$); $p(t)$ – интервал простоя признака классификации на текущий момент времени ($p_{\max} \geq p(t) \geq 0$); p_{\max} – максимально возможный (допустимый) интервал простоя признака классификации.

При условии $p(t) = p_{\max}$ признак классификации перестает быть актуальным и может быть удален из системы фильтрации в зависимости от требований метасистемы.

Основной характеристикой признаков классификации служит их достоверность (коэффициент достоверности). Коэффициент достоверности признака идентификации – скалярная величина, показывающая долю правильно классифицированных сообщений к общему количеству отобранных материалов по заданному признаку:

$$\alpha(t) = \frac{v_{\text{клас}}(t)}{v(t)}, \quad (8)$$

где $\alpha(t)$ – коэффициент достоверности признака классификации в текущий момент времени ($1 \geq \alpha(t) \geq 0$); $v_{\text{клас}}(t)$ – количество правильно классифицированных сообщений по заданному признаку на текущий момент времени; $v(t)$ – общее количество сообщений, классифицированных по заданному признаку в текущий момент времени.

Границы диффузии – [0; 1]. При условии наличия в тракте обработки только сообщений, представляющих интерес для конечного пользователя информации, и максимального качества классификации – диффузия максимальна и равна 1. В случае отсутствия сообщений в тракте обработки диффузия минимальна и равна 0.

Уравнение (3) описывает диффузию материалов в тракте обработки с учетом доли сообщений представляющих интерес для потребителей информации, информативности классификационных признаков, а также предела изменений, согласно пропускной способности системы обработки, в конкретный момент времени.

Таким образом, математическая модель процесса фильтрации сообщений на основе диффузии материалов в тракте обработки системы фильтрации входной информации имеет вид

$$\frac{dx}{dt} = \frac{N(t)}{Y(t)} \cdot \sum_{i=1}^n \frac{x_i(t)}{N(t)} \cdot \frac{\sum_{j=1}^M \frac{\sum_{k=1}^{p_{\max}} \frac{v_{\text{клас } j, k}(t-k)}{v_j(t-k)} \cdot (1 - \frac{p_j(t)}{p_{\max j}})}{k}}{M_{\text{при}}(t)}. \quad (9)$$

В результате проведения имитационного моделирования было установлено, что рассматриваемая математическая модель адекватно отражает особенности ведения технологического процесса отбора сообщений, представляющих интерес для конечных потребителей, обусловленных пропускной способностью системы фильтрации и множеством классификационных признаков. В таблице представлены значения параметров, используемых при моделировании.

В ходе исследования было сформировано множество актуальных и достоверных классификационных признаков – 231, позволяющее вести процесс фильтрации информации наиболее эффективно, т.е. максимизировать поток сообщений, представляющих интерес для конечных пользователей, в тракте обработки системы фильтрации (рис. 2).

Значения параметров при моделировании

Параметр	Значение
Входной поток сообщений	13 000
Количество сообщений, представляющих интерес	1 382
Пропускная способность системы фильтрации	15 000
Начальное количество классификационных признаков	768

Относительно малые значения диффузии объясняются пропускной способностью системы фильтрации и количеством интересующих сообщений в тракте обработки – 9,21% от максимально возможного потока. При увеличении потока интересующих сообщений (или уменьшении пропускной способности) и множестве классификационных признаков, обеспечивающих эффективную фильтрацию, значения диффузии возрастают и стремятся к максимуму.

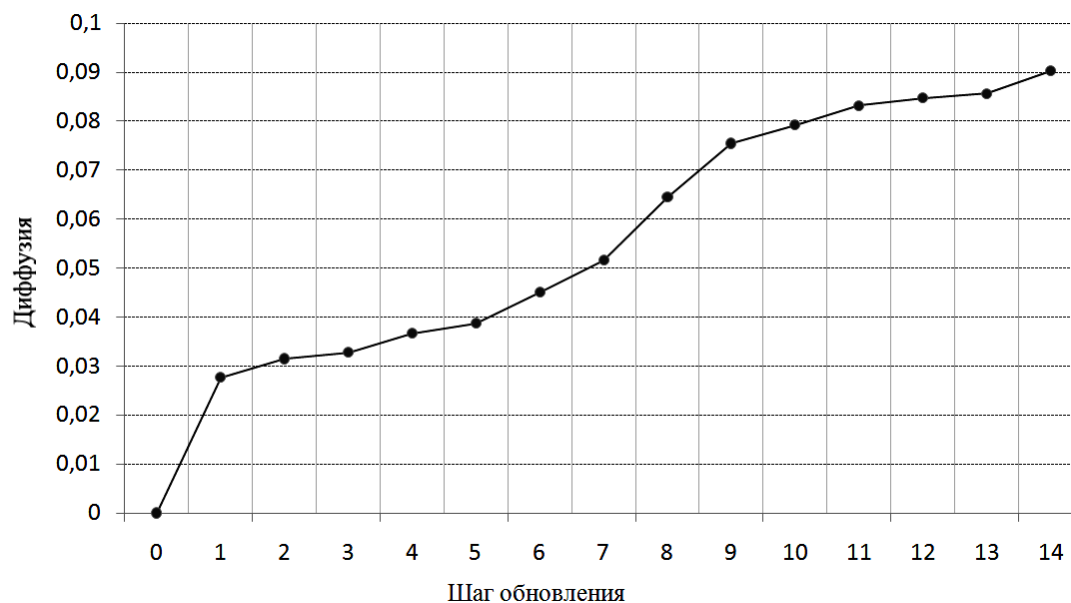


Рис. 2. Диффузия сообщений, представляющих интерес, в тракте обработки системы фильтрации

Представленная математическая модель процесса фильтрации сообщений на основе диффузии материалов в тракте обработки может рассматриваться как показатель эффективности функционирования системы фильтрации информации, учитывающей особенности предметной области, – пропускная способность системы и множество классификационных признаков.

Литература

1. Семененко М.Г. Введение в математическое моделирование. – М.: Солон-Р, 2002. – 112 с.
2. Самарский А.А. Математическое моделирование. Идеи. Методы. Примеры / А.А. Самарский, А.П. Михайлов. – М.: Физмат, 2005. – 320 с.
3. Абланская Л.В. Экономико-математическое моделирование / Л.В. Абланская, Л.О. Бабешко, Л.И. Баусов. – М.: Экзамен, 2006. – 800 с.
4. Яблонский А.И. Математические модели в исследовании науки. – М.: Эдиториал УРСС, 2001. – 400 с.
5. ГОСТ 3.1109-82. Единая система технологической документации. Термины и определения основных понятий. – М.: Изд-во стандартов, 1983. – 15 с.
6. Дворцин М.Д. Технодинамика: Основы теории формирования и развития технологических систем / М.Д. Дворцин, В.Н. Юсим. – М.: Международный фонд истории науки «Дикси», 1993. – 320 с.
7. Московкин В.М. К построению математических основ концепции диффузии инноваций // Сборник науч. трудов Харьковского института социального прогресса. – Харьков: Бизнес-информ, 1997. – № 2. – С. 109–112.

8. Гурков И.Б. Инновации в российской промышленности: создание, диффузия и реализация новых технологий и социальных практик / И.Б. Гурков, В.С. Тубалов // Мир России. – 2004. – Т. 13, № 3. – С. 28–47.
9. Щепина И.Н. Модели диффузии и замещения нововведений (обзор литературы). – М.: ЦЭМИ РАН, 1990. – 48 с.
10. Серков Л.А. Синергетические аспекты моделирования социально-экономических процессов. – Екатеринбург: ИЭ УрО РАН, 2008. – 214 с.
11. Франк-Каменский Д.А. Диффузия и теплопередача в химической кинетике. – М.: Наука, 1987. – 502 с.
12. Гнеледорф П. Термодинамическая теория структуры, устойчивости и флуктуаций / П. Гнеледорф, И. Пригожин. – М.: Мир, 1973. – 280 с.
13. Николис Г. Самоорганизация в неравновесных системах. От диссипативных структур к упорядоченности через флуктуации / Г. Николис, И. Пригожин. – М.: Мир, 1979. – 512 с.
14. Москвин В.М. Основы концепции диффузии инноваций // Бизнесинформ (Харьков). – 1998. – №17. – 12 с.
15. Ту Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М.: Мир, 1978. – 413 с.
16. Горяев Ю.А. Информатика: учеб. пособие. – М.: МИЭМП, 2007. – 116 с.

Родин Евгений Александрович
Инженер ФГУП «НИИ Квант», г. Москва
Тел.: +7-903-565-58-20
Эл. почта: earodin@rambler.ru

Rodin E.A.

Mathematical model of selection process of messages on the basis of material diffusion in the system of information filtrations

In the automated systems of information filtration there is a problem of estimation of quality of its functioning, connected with the complexity of classification of an end-product – information, and the uncertainty concerning information sources. The paper considers the mathematical model of selection process of messages on the basis of material diffusion in the processing path, considering the peculiarities of filtration system – set of classification signs and carrying capacity of a filtration system.

Keywords: system of messages filtration, diffusion, material diffusion in the processing path, information classification.