

УДК 004.8:004.49

С. С. ВАЛЕЕВ, А. П. НИКИТИН

МНОГОУРОВНЕВАЯ СИСТЕМА ФИЛЬТРАЦИИ СПАМА НА ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Рассматривается задача автоматической фильтрации электронных сообщений. Предлагается многоагентная многоуровневая архитектура системы фильтрации спама. Ставится и решается задача классификации сообщений на основе когнитивного представления информации и линейного нейросетевого классификатора. Многоагентная программная система; фильтрация сообщений; нейросетевой линейный классификатор

ВВЕДЕНИЕ

Электронные сообщения рекламного характера (спам), отправляемые автоматически большому количеству адресатов, занимают до 90% от общего почтового трафика [1–3]. Программные средства фильтрации спама (спам-фильтры), устанавливаемые на почтовых серверах или на компьютерах адресатов, решают задачу классификации (категоризации) текстов. Задача классификации сообщений для заданного множества сообщений $M = \{m_i, i = 1 \div n\}$ и множества заданных классов $C = \{c_j, j = 1 \div k\}$ заключается в поиске отображения $K = M \times C \rightarrow \{1, 0\}$, такого, что $K(m_i, c_j) = 1$ при $m_i \in c_j$ и $K(m_i, c_j) = 0$ при $m_i \notin c_j$. Отображение K , если оно существует, называется бинарным классификатором.

При фильтрации спама наиболее широкое распространение получили байесовские классификаторы [4–7]. Показано, что при определенных допущениях байесовский классификатор является линейным классификатором вида $y = w^T x + b$, где x — вектор наблюдения, w — вектор весов, b — скалярная величина (порог) [8]. При принадлежности вектора наблюдений обеим категориям сообщений возникает ситуация неопределенности.

Для обучения спам-фильтров используются база данных спама (БД спама) и база легитимных сообщений (БД легитимных сообщений) [6, 7]. На основе этой выборки формируются частотные словари ключевых лексем (слов). В зависимости от метода фильтрации они могут включать до нескольких тысяч лек-

сем [7]. Также могут учитываться особенности морфологии языка сообщений.

В [9–10] приводятся описания классификаторов, построенных на базе нейронных сетей и нечеткой логики, учитывающих частотные свойства базовых лексем сообщений.

К основному недостатку известных спам-фильтров следует отнести разрушение семантической составляющей сообщений при формировании обучающей выборки и категоризации сообщений.

Следует отметить, что качество решения задачи фильтрации сообщений в том числе определяется архитектурой системы фильтрации. При централизованной фильтрации (т.е. спам-фильтр установлен на почтовом сервере организации) проявляется эффект усреднения правил фильтрации, что может привести к потере важного для адресата сообщения. При фильтрации спама на компьютере конечного пользователя возрастает нагрузка на сетевую инфраструктуру организации. В связи с этим ставится задача разработки системы фильтрации спама, обеспечивающей централизованную фильтрацию спама с учетом предпочтений всех конечных пользователей сети и с учетом семантической составляющей сообщений.

1. МНОГОУРОВНЕВАЯ СИСТЕМА ФИЛЬТРАЦИИ СПАМА

В качестве архитектуры системы фильтрации спама предлагается использовать иерархическую организацию процесса обработки информации. На верхнем уровне (уровне централизованной фильтрации) собирается ин-

формация о результатах категоризации сообщений на каждой рабочей станции. В результате при фильтрации сообщений можно учитывать не только правила фильтрации, заданные администратором почтового сервера, но и предпочтения конечного пользователя. На рис. 1, а схематично представлен процесс фильтрации при традиционном способе построения спам-фильтра, а на рис. 1, б представлен процесс многоуровневой обработки сообщений, что позволяет снизить нагрузку на сетевую инфраструктуру организации.

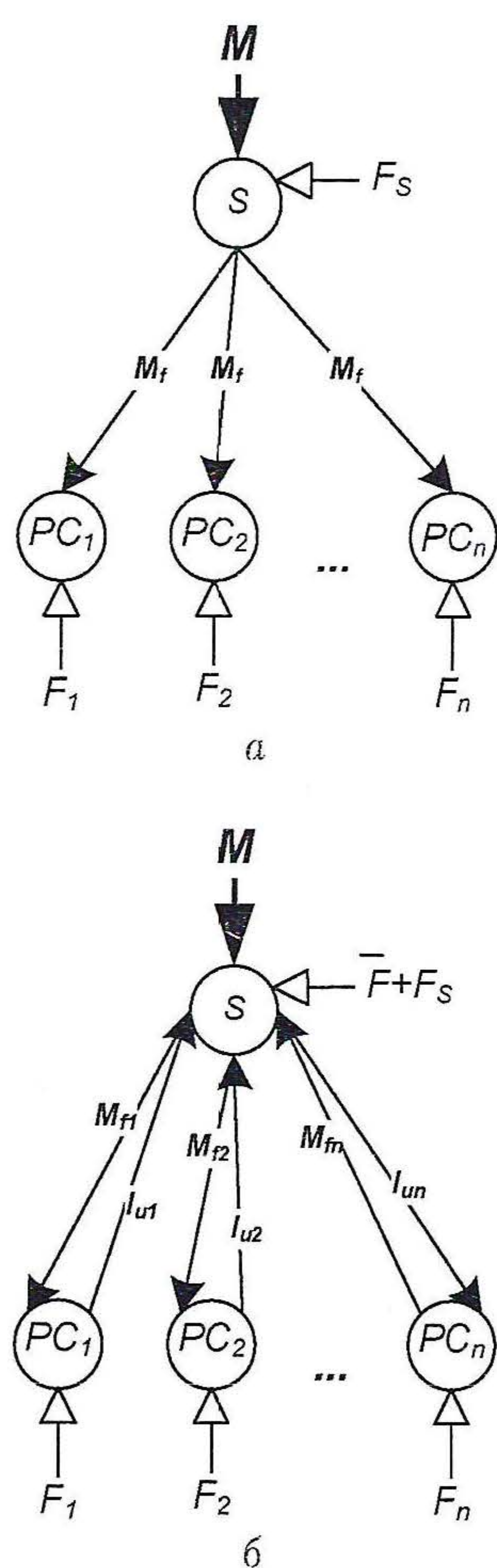


Рис. 1. Процессы фильтрации спама: M — поток входящих сообщений; S — почтовый сервер; M_f — поток отфильтрованных сообщений; F_s — правила фильтрации, заданные администратором почтового сервера; F_i — правила фильтрации, заданные i -м пользователем ($i = 1 \div n$); PC_i — рабочая станция i -го пользователя ($i = 1 \div n$); M_{fi} — поток сообщений, отфильтрованных с учетом правил F_s и правил F_i ; I_{ui} — результаты категоризации сообщений на i -й рабочей станции ($i = 1 \div n$)

В качестве основы построения предлагаемой многоуровневой системы фильтрации спама использован многоагентный подход. На рис. 2 представлена архитектура прототи-

па многоагентной системы фильтрации спама.

Особенностью данной многоагентной системы фильтрации спама является асинхронная обработка почтовых сообщений и наличие обратной связи с пользовательских компьютеров, заключающаяся в возможности дообработки нераспознанных сообщений на клиентском компьютере и связи с сервером базы данных (БД) сообщений. Для учета особенностей предпочтений конечных пользователей используется двухуровневая организация процесса фильтрации. Основная задача верхнего уровня — это организация процедур обучения классификаторов на конечных компьютерах пользователей. В результате обучения на каждом j -м компьютере для i -го сообщения формируется вектор атрибутов $k_j^i = \{s, l\}$, где s — атрибут спама, l — атрибут легитимного сообщения, $s, l \in \{0, 1\}$, $i = 1 \div k$, $j = 1 \div n$.

Прототип предлагаемой системы реализован на платформе JADE (Java Agent Development Framework) [11]. Особенностью данной платформы является возможность реализации поддержки распределенной обработки информации на базе многоагентной технологии, поддержки архитектуры P2P, использование языка ACL (Agent Communication Language), удовлетворяющего требованиям FIPA (Foundation for Intelligent Physical Agents).

2. НЕЙРОСЕТЕВОЙ КЛАССИФИКАТОР СООБЩЕНИЙ

Как отмечалось ранее, при решении задачи классификации спама в рамках известных подходов используется синтаксический анализ: текст d_i разбивается на лексемы, которые затем используются в качестве атрибутов при классификации. Далее с их помощью принимается решение о принадлежности сообщения к какой-то категории. Недостатком таких фильтров является невозможность учета семантической составляющей электронных сообщений.

Разработаны системы извлечения знаний из больших массивов данных на основе алгоритмов DataMining, позволяющие извлекать семантическую компоненту, которые могут быть использованы при построении систем интеллектуальной фильтрации спама [12]. Однако, как известно, они отличаются высокой сложностью, что затрудняет их использование для решения задачи фильтрации спама как на конечном компьютере пользователя, так и на сервере почтовых сообщений.

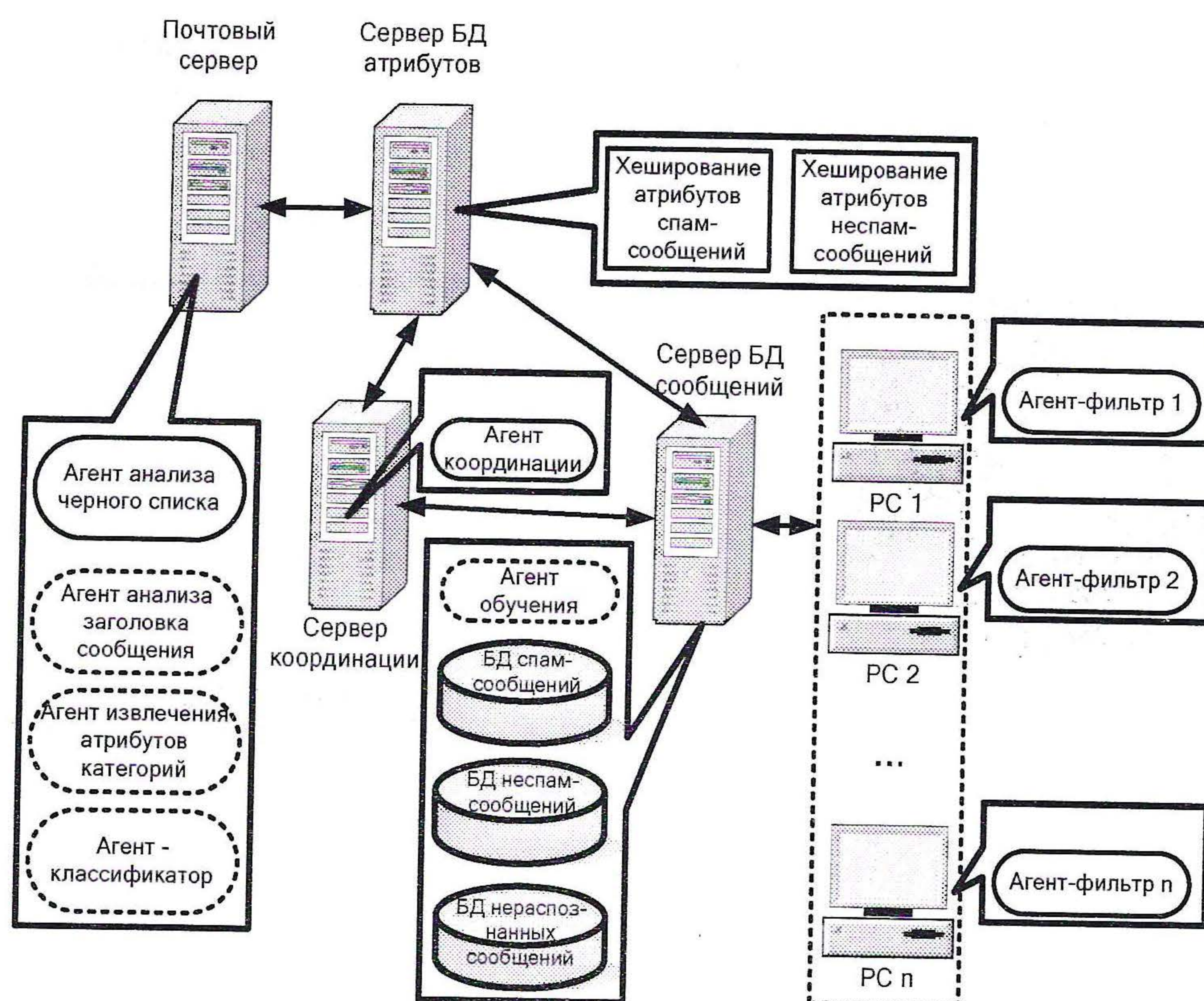
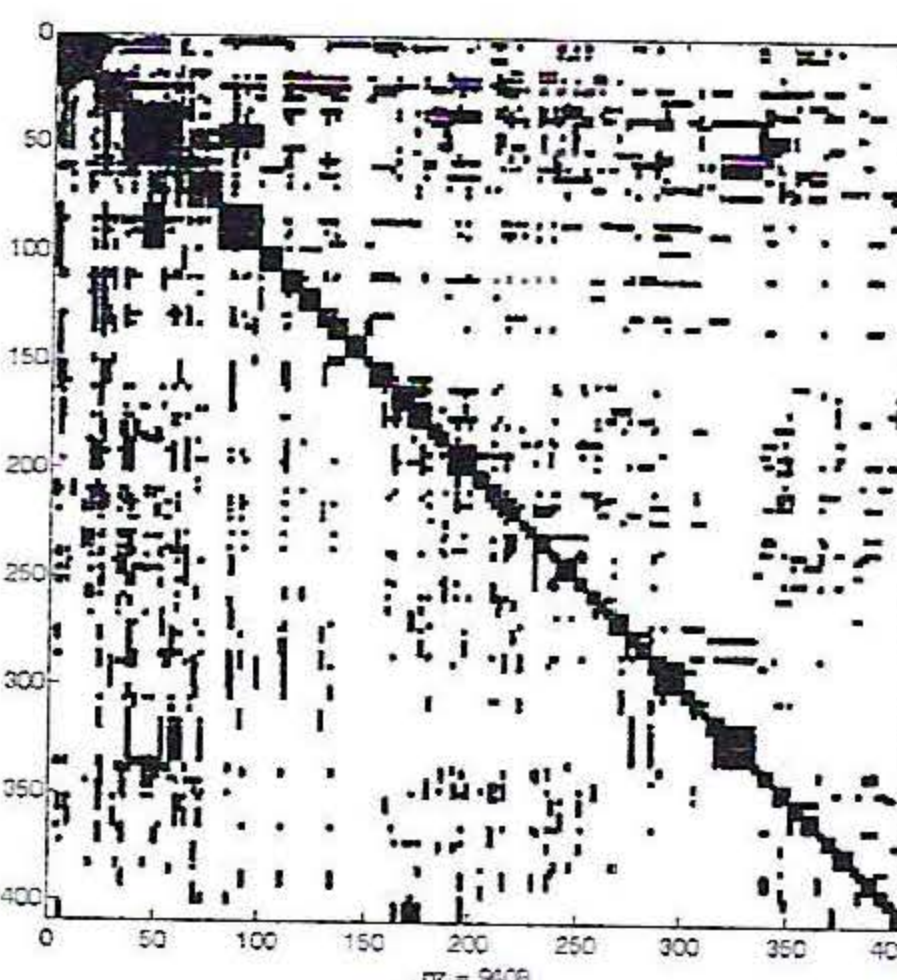


Рис. 2. Многоагентная система фильтрации спама

В рамках предлагаемого метода при решении задачи построения интеллектуального классификатора предлагается не только оценивать частоту появления лексем в сообщении, но и определять меру близости лексем в рамках принятой минимальной семантической структуры сообщения. Минимальной семантической структурой в текстовых сообщениях можно, например, принять предложение. С учетом небольшого объема электронных писем это позволяет извлекать часть семантической информации, которая может быть использована в процессе категоризации сообщений.

Рассмотрим основные этапы процесса категоризации сообщений на основе предлагаемого подхода. На первом этапе агентом-классификатором из полученного текстового сообщения формируется множество $F = \{f_k\}$, где f_k — заданная для системы фильтрации минимальная семантическая структура, $i = 1 \div m$, при этом из структуры удаляются избыточные элементы, но сохраняется последовательность лексем. В результате структура f_k может быть представлена в виде семантического графа s_k , определяющего связь лексем в f_k . При оценке силы связи учитывается лексикографический порядок в f_k . После обработки всего сообщения получаем семантический граф $G = \langle S, V \rangle = \bigcup_1^n s_k$,

представляющий собой упрощенную семантическую модель сообщения, где S — множество вершин, содержащее лексем сообщения, прошедшие предварительную фильтрацию на основе заданного словаря; V — множество ребер, весовые значения которых определяют силу связи лексем в f_k . Граф G задается матрицей смежности Z (семантическая матрица). На рис.3 представлена в графическом виде матрица смежности $Z \in R^{400 \times 400}$ для 200 сообщений, полученных по электронной почте. Как следует из рис. 3, матрица Z имеет четко выраженную структуру, зависящую от семантической структуры сообщений, и тем самым эта информация может быть использована для повышения эффективности процедуры классификации сообщения.

Рис. 3. Графическое представление матрицы смежности семантического графа G

На следующем этапе фильтрации выполняется понижение размерности матрицы Z (сжатие образа), из нее удаляются столбцы и строки, содержащие значения элементов ниже заданного порога λ (рис. 4).

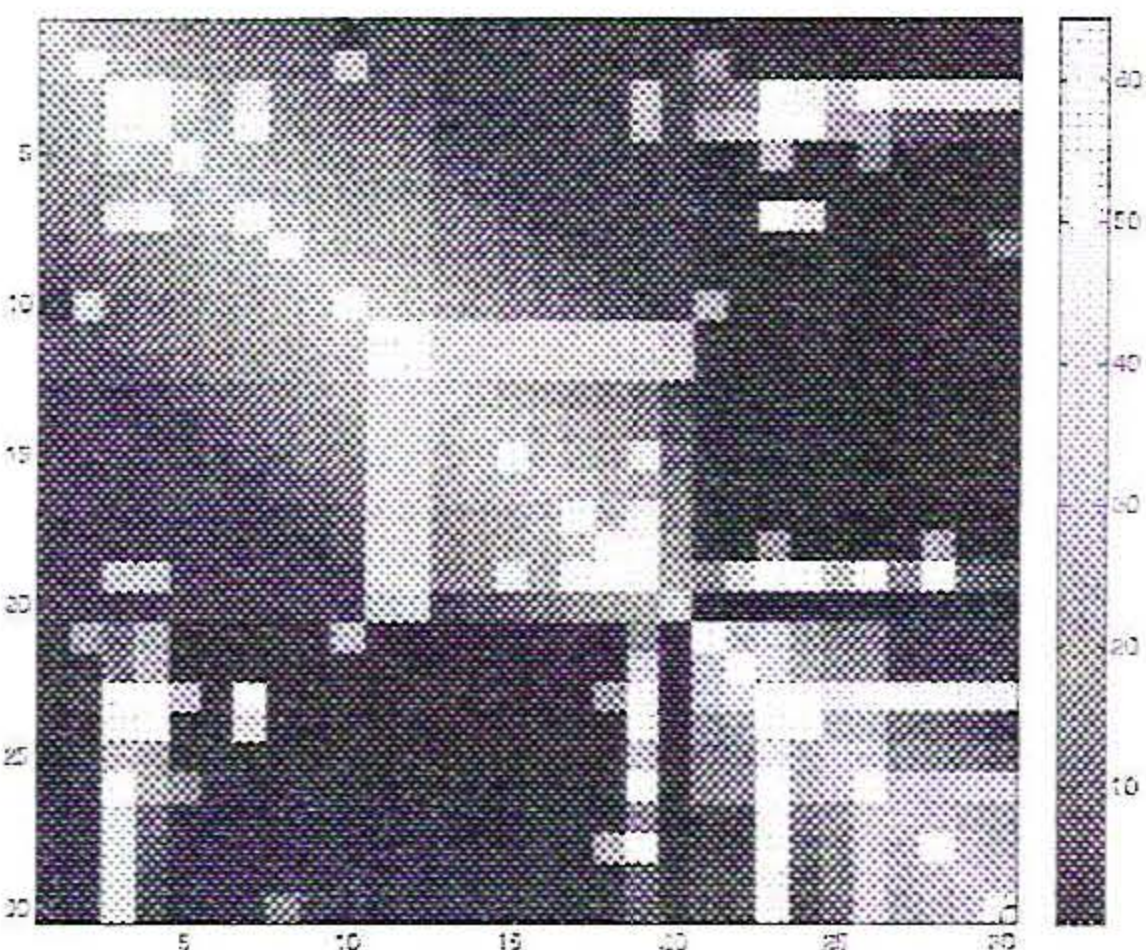


Рис. 4. Графическое представление матрицы G после ее сжатия

На рис. 4 представлена шкала весов (сила связей между лексемами в семантическом графе).

Далее с учетом векторов атрибутов сообщений на базе G строятся две обобщенные семантические матрицы Z_s (для спама) и Z_e (для легитимных сообщений). В результате задача классификации новых сообщений сводится к построению бинарного классификатора N , определяющего принадлежность сообщения к одной из представленных семантическими матрицами категорий.

В качестве классификатора выбран линейный нейросетевой ассоциатор вида $y = w^T x$, где x — входной вектор (элементы вектора составлены из элементов семантической матрицы фильтруемого сообщения), w — вектор весов классификатора (рис. 5, а, где R — размерность вектора x). Обучение ассоциатора выполняется на основе правила обучения Хебба [8]. К достоинству этого классификатора следует отнести простоту его обучения, что является определяющим качеством при выборе его в качестве ядра системы интеллектуальной фильтрации спама. На рис. 5, б представлен интерфейс классификатора, реализованный на базе языка C# и позволяющий настраивать классификатор на начальном этапе обучения системы фильтрации спама. Архитектура разработанной системы многоуровневой фильтрации спама позволяет реализовать способность к самообучению. На рис. 6 представлена диаграмма изменения доли спама и легитимных сообщений в процессе обучения многоагентной системы фильтрации.

Как следует из диаграммы, доля спама в сообщениях, получаемых конечным пользователем, сократилась в 6 раз. Тем самым под-

тверждается достаточно высокая эффективность предлагаемого подхода.

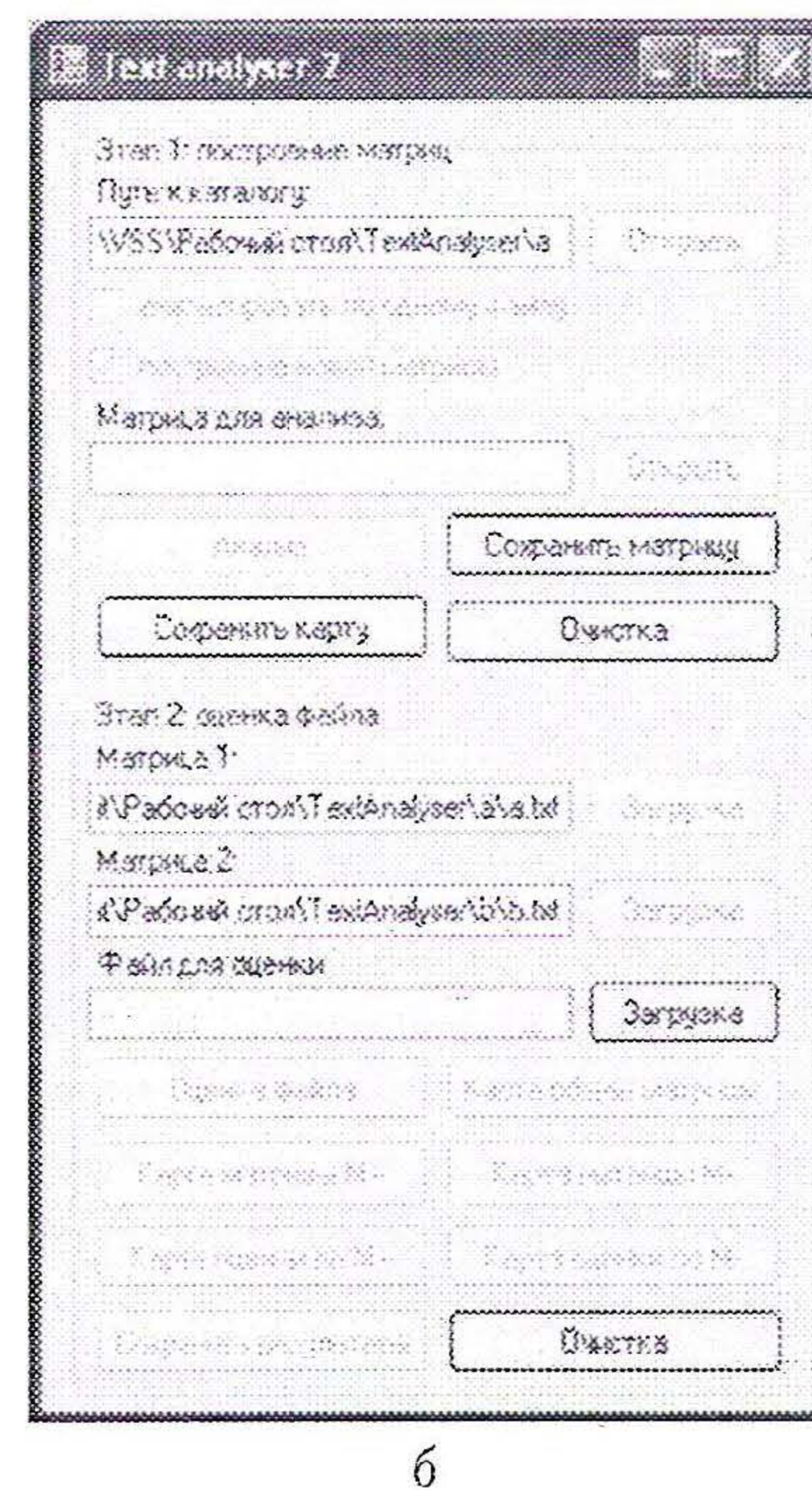
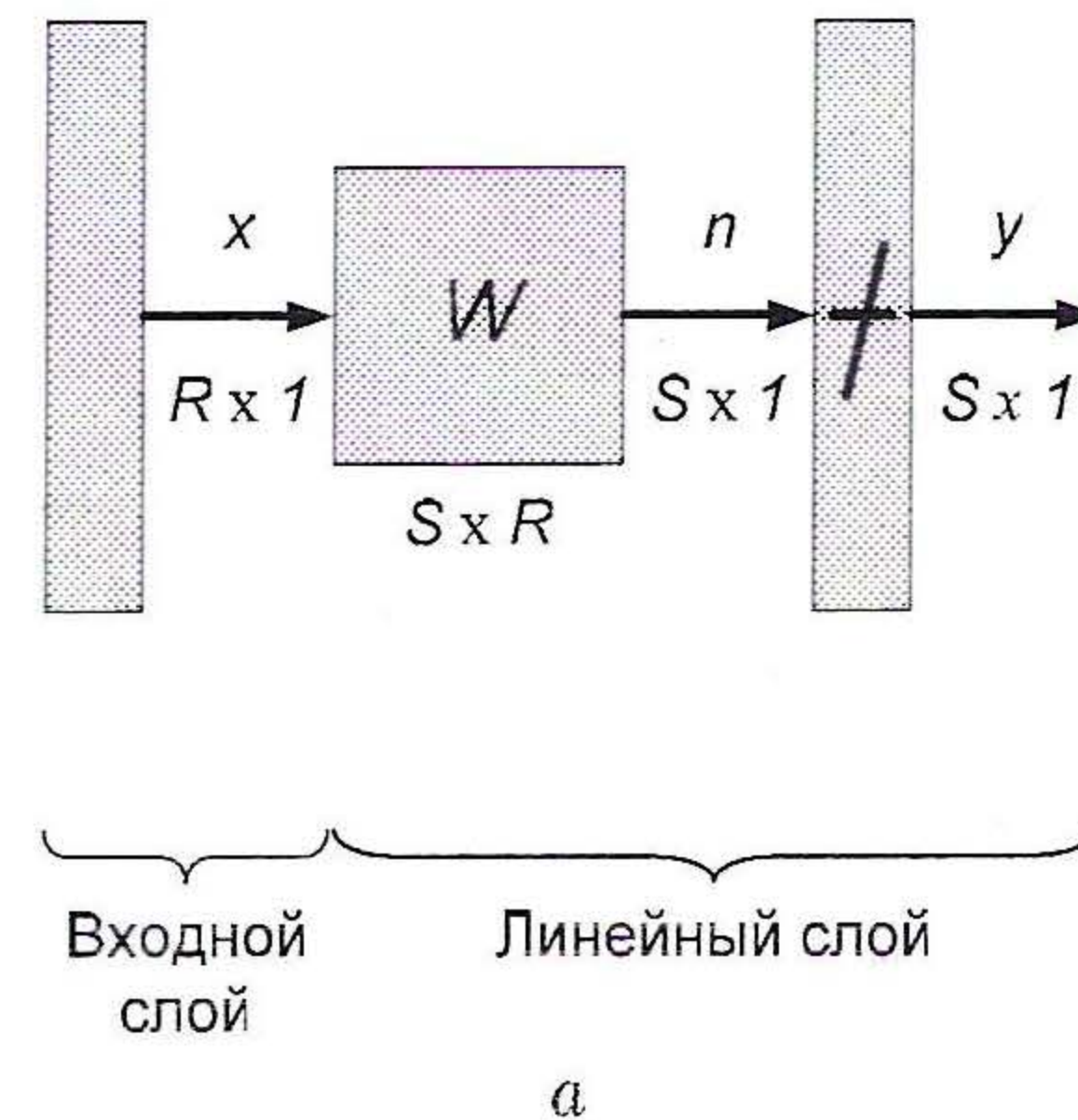


Рис. 5. Линейный нейросетевой классификатор



Рис. 6. Диаграмма изменения доли спама и легитимных сообщений в почтовом трафике

ЗАКЛЮЧЕНИЕ

Многоуровневая многоагентная система фильтрации спама на базе иерархической системы классификации сообщений позволяет учитывать предпочтения конечных пользователей при фильтрации спама на почтовом сервере. Обучение классификатора сообщений основано на использовании упрощенного семантического графа сообщения в качестве элемента обучающей выборки, что позволяет свести задачу фильтрации к решению задачи распознавания образов и повысить качество автоматической классификации сообщений. Предлагаемая система реализована на многоагентной платформе JADE, ориентированной на язык Java, что обеспечивает переносимость системы, а также возможность ее масштабирования.

СПИСОК ЛИТЕРАТУРЫ

1. **Дронов, В.** Портрет заказчика спамерских услуг в России / В. Дронов [Электронный ресурс] (<http://www.spamtest.ru/document?pubid=190000936&context=1>).
2. **Gorev, A.** Методы борьбы со спамом / A. Gorev [Электронный ресурс] (http://www.opennet.ru/base/net/spam_greylist.txt.html).
3. **Власова, А.** Спам как рекламный инструмент / А. Власова [Электронный ресурс] (<http://www.spamtest.ru/document?pubid=193501834&context=1>).
4. **Sahami, M.** A bayesian approach to filtering junk e-mail / M.Sahami, S. Dumais, D. Heckerman, E. Horvitz // Learning for Text Categorization, Wisconsin, 1998.
5. **Androutsopoulos, I.** An Evaluation of Naive Bayesian Anti-Spam Filtering / I. Androutsopoulos, J. Koutsias, K. Chandrinos, G. Paliouras // Proc. of the 11th European Conference on Machine Learning (ECML 2000), Barcelona, 2000. P. 9–17.
6. **Graham, P.** A plan for spam / P. Graham. [Электронный ресурс] (<http://www.paulgraham.com/spam.html>).

7. **Graham, P.** Better bayesian filtering / P. Graham // Proc. of the 1st Annual Spam Conference, MIT, 2003. [Электронный ресурс] (<http://www.paulgraham.com/better.html>).
8. **Хайкин, С.** Нейронные сети: полный курс / С. Хайкин. М.: Вильямс, 2005. 1104 с.
9. **Stuart, I.** A Neural Network Classifier for Junk E-Mail / I. Stuart, S. Cha, C. Tappert // DAS. Lecture Notes in Computer Science. Vol. 3163. Springer-Verlag, 2004. P. 442–450.
10. **Kim, J.** Determination of Usenet News Groups by Fuzzy Inference and Kohonen Network / J. Kim, H. Kim, S. Kang, B. Kim // Lecture Notes in Artificial Intelligence. Vol. 3157. Springer-Verlag, 2004. P. 654–663.
11. **Bellifemine, F.** Java Agent Development Framework an Open Source platform for peer-to-peer agent based applications / F. Bellifemine. [Электронный ресурс] (<http://jade.tilab.com>).
12. **Hand, D. J.** Principles of Data Mining / D. J. Hand, H. Mannila, P. Smyth. The MIT Press. Cambridge, MA, 2001. 546 p.

ОБ АВТОРАХ



Валеев Сагит Сабитович, проф. каф. вычисл. техн. и защиты информации. Дипл. инж.-электромех. (УАИ, 1970). Д-р техн. наук по упр. в техн. системах (УГАТУ, 2005). Иссл. в обл. интеллект. упр. сложными объектами.



Никитин Андрей Павлович, асп. той же каф. Дипл. спец. по защите информации (УГАТУ, 2005). Иссл. в обл. информ. безопасности и интеллект. систем.