

**Московский Государственный Технический Университет
им. Н.Э. Баумана**

Цыганов Илья Германович

***МНОГОАГЕНТНАЯ АВТОМАТИЗИРОВАННАЯ СИСТЕМА АДАПТИВНОЙ ФИЛЬТРАЦИИ
ПОТОКОВ ТЕКСТОВОЙ ИНФОРМАЦИИ***

**Специальность 05.13.01 «Системный анализ, управление и обработка информации (в
науке и промышленности)»**

А в т о р е ф е р а т

**диссертации на соискание ученой степени
кандидата технических наук**

Москва - 2004

Работа выполнена в Московском Государственном Техническом Университете
им. Н.Э. Баумана

Научный руководитель

канд. техн. наук, доцент
Власов Андрей Игоревич

Официальные оппоненты:

Ведущая организация

Федеральное государственное унитарное предприятие
«Научно-производственное предприятие «Пульсар»
(Москва)

Защита диссертации состоится « ____ » _____ 2010 г. в ____ часов на заседании диссертационного совета _____ в Московском государственном техническом университете им. Н.Э. Баумана по адресу: 105005, г. Москва, 2-ая Бауманская ул., д.5, стр.1.

С диссертацией можно ознакомиться в библиотеке Московского государственного технического университета им. Н.Э. Баумана.

Ваш отзыв в 1 экземпляре просим высылать по указанному адресу.

Автореферат разослан « ____ » _____ 2010 г.

Телефон для справок: 8(499) 267-89-63.

Ученый секретарь диссертационного совета
д.т.н., профессор

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность:

Автоматические системы фильтрации (АСФ) информационных потоков представляют собой обширный класс систем, использующихся для защиты пользователей от получения информации нежелательного характера. АСФ широко используются в системах электронной почты глобальной сети Интернет, где угрожающих масштабов достигла незапрашиваемая рассылки (НР) сообщений коммерческого характера. Однако в ответ на широкое внедрение АСФ НР, распространители НР стали активно применять различные методы, позволяющие обойти автоматические алгоритмы фильтрации, что значительно повысило уровень ошибок АСФ НР и снизило их эффективность. Поэтому сейчас возникла потребность в разработке новых подходов к фильтрации, основанных на принципах адаптации при выявлении ошибок.

Во всех современных АСФ НР, однако, настройка фильтров при выявлении ошибок до сих пор осуществляется в ручном режиме и требует постоянного участия оператора, обладающего специальными знаниями и полномочиями. В современных условиях, когда объемы НР существенно возросли, использование ручного труда в контуре обратной связи приводит к значительным задержкам и/или росту затрат на эксплуатацию системы. Поэтому сейчас актуальной является задача построения гибкой адаптивной системы фильтрации, способной к самонастройке, осуществляемой в автоматическом режиме при получении информации об ошибках фильтрации, поступающей от пользователей систем электронной почты в реальном масштабе времени.

Решение поставленной задачи включает два основных аспекта: 1) разработка архитектуры системы, позволяющей учитывать голоса пользователей при выявлении новых видов НР; 2) разработка адаптивного алгоритма, позволяющего автоматически настраивать систему фильтрации по выявленным примерам сообщений НР.

В настоящее время корпоративные системы электронной почты являются одними из наиболее сложных в наладке и эксплуатации. Переход на новый тип систем связан с необходимостью переобучения административного персонала и пользователей, что неизбежно приводит к убыткам предприятия. Поэтому сейчас актуальным является разработка АСФ НР на базе существующей инфраструктуры и стандартного программного обеспечения (ПО) систем электронной почты. Новые возможности АСФ НР должны гибко интегрироваться в существующие системы, не изменяя привычные для пользователя функции.

В условиях, когда от пользователя поступает информация о том, какие из полученных сообщений являются НР, адаптация фильтров основывается на выделении характерных для НР признаков (слов и словосочетаний), содержащихся в текстах НР. Основные сложности в таком анализе связаны с тем, что сообщения НР имеют различную тематику, поэтому такой анализ предполагает выделение в текстах тематических групп и характерных для них подмножеств слов и словосочетаний. Подобная задача в условиях, когда размерность велика, не может быть решена с помощью классических линейных методов. Поэтому актуальным является разработка нелинейных методов адаптации. В работе для этого используются последние достижения в области построения нелинейных нейросетевых классификаторов с переменной структурой.

Большой вклад в развитие теории нейронных сетей и ее применении к решению задач адаптивной фильтрации и классификации внесли Хебб Д., Минский М., Розенблат Ф., Уидроу Б., Андерсон, Кохонен Т., Гроссберг С., Коско Б., Галушкин А.И., Горбань А.Н., Мкртчян С.О. и многие другие.

Разрабатываемые в работе принципы, модели, методы и алгоритмы могут быть использованы во многих прикладных областях, например, в системах управления документами (СУД); системах управления деловыми процессами (СУДП); системах управления внешними и внутренними потоками информации (например, на базе Web) в системах передачи данных (СПД) предприятий; системах адаптации к профилю пользователя; системах обеспечения заданного уровня безопасности и многих других.

Цель работы. Цель работы состоит в разработке средств фильтрации потоков нового поколения, способных повысить эффективность функционирования существующих АСФ НР за счет:

- 1) создания эффективных схем учета и использования информации об ошибках фильтрации, поступающей от пользователей;
- 2) привлечения пользователей систем электронной почты к настройке фильтров;
- 3) минимизации и автоматизации ручного труда при настройке фильтров;
- 4) реализации алгоритмов адаптации фильтров в реальном масштабе времени при появлении уже первой информации об ошибках фильтрации;
- 5) использования современных нелинейных нейросетевых методов синтеза и настройки фильтров;
- 6) расширения функциональных и сервисных возможностей современных корпоративных АСФ НР.

Для достижения поставленной цели в работе был решен комплекс задач, в частности:

- 1) анализ возможностей существующих корпоративных АСФ НР;
- 2) определение требований к функциональному составу, алгоритмическому и программному обеспечению АСФ НР.
- 3) разработка и исследование многоагентной архитектуры АСФ НР, позволяющей пользователям участвовать в процессах фильтрации и адаптации;
- 4) разработка комплекса средств, позволяющих производить сквозной учет поступающих в корпоративную систему сообщений;
- 5) разработка математических методов, позволяющих производить формирование обучающей выборки с учетом голосов пользователей по отдельным сообщениям;
- 6) разработка математических методов выбора пространства признаков в задаче анализа содержания сообщений;
- 7) разработка адаптивных алгоритмов синтеза нейронной сети с переменной структурой, позволяющих осуществлять фильтрацию НР в реальном масштабе времени;
- 8) исследование и выбор методов программной и аппаратной реализации средств фильтрации НР;
- 9) разработка аппаратно-программного комплекса фильтрации НР на основе разработанных принципов;
- 10) экспериментальное исследование, подтверждающее эффективность предложенных моделей, методов и алгоритмов в реальных условиях.

Методы исследования. При решении поставленных задач использованы элементы теории информационных систем, передачи и обработки данных, элементы теории экспертных систем и обработки знаний, нейроматематики, теории нейронных сетей, теории оптимизации, математический аппарат теории автоматического управления, теории вероятностей и математической статистики, линейной алгебры, теории марковских случайных полей, элементы теории эргодичности, положения теории автоматизации проектирования, теории конструирования и надежности ЭА.

Научная новизна работы состоит в следующем:

- 1) Проведен анализ и систематизация возможностей существующих корпоративных АСФ НР и методов их интеграции в системы электронной почты.
- 2) Выработаны требования к функциональному составу, алгоритмическому и программному обеспечению, выполнение которых необходимо для построения эффективных в современных условиях АСФ НР.
- 3) На базе стандартной инфраструктуры систем электронной почты разработана и исследована многоагентная архитектура АСФ НР, позволяющая отдельным пользователям участвовать в процессе фильтрации НР: воздействовать на алгоритмы настройки фильтров и осуществлять выбор степени жесткости фильтрации в индивидуальном порядке.

- 4) Разработан комплекс средств, позволяющих производить сквозной учет поступающих в корпоративную систему сообщений.
- 5) Разработаны математические методы, позволяющие производить формирование обучающей выборки с учетом значимости голосов отдельных пользователей по поступившим в систему сообщениям.
- 6) Разработаны математические методы выбора пространства признаков в задаче анализа содержания сообщений, что позволило повысить точность анализа за счет учета значимых словосочетаний.
- 7) Разработан метод синтеза нейронной сети с переменной структурой, входным сигналом которой являются разряженные векторы большой размерности (десятки тысяч).
- 8) Проведены исследования и выбор методов программной и аппаратной реализации средств фильтрации НР.
- 9) Разработан аппаратно-программный комплекс фильтрации НР на основе разработанных принципов;

Достоверность полученных научных результатов, выводов и рекомендаций диссертационной работы подтверждена результатами экспериментальных исследований, проведенных на базе почтового портала концерна “РТИ-Системы”. Полученные в работе результаты наглядно демонстрируют эффективность использования разработанных моделей, методов и алгоритмов для решения задач автоматической фильтрации незапрашиваемой рассылки.

В период с 2004 г. по 2005 г. проведены работы по внедрению разработанной АСФ НР в корпоративной системе электронной почты Международного Института Экономики и Права.

Естественно, что в условиях, когда пользователи индивидуально регулируют степень жесткости фильтрации, оценка качества системы существенно зависит от выбираемых параметров. Во время эксплуатации данной системы на базе почтового портала концерна “РТИ-Системы” были проведены работы по построению трех пользовательских профилей системы: “жесткая”, “умеренная” и “мягкая” фильтрация. Средняя вероятность правильной фильтрации сообщений НР соответственно приведенным профилям составила: 68,2%, 79,6%, 82,0%. Вероятность ложной фильтрации пользовательских сообщений: 0%, 0,01%, 1,4%. Для сравнения, современные коммерческие системы в режиме с уровнем ложных сработок порядка 2-3% достигают пикового показателя вероятности правильной фильтрации 70-75%.

Положения, выносимые на защиту:

- 1) Архитектура, функциональный состав и интерфейсы АСФ НР, реализованные, позволяют.....
- 2) Система сквозного учета сообщений, позволяет управлять системой фильтров, что обеспечивает
- 3) Метод формирования обучающей выборки с учетом голосов пользователей по отдельным сообщениям и уровня доверия к каждому пользователю, обеспечивает
- 4) Метод выбора пространства признаков в задаче контекстной обработки потоков текстовой информации позволяет учитывать значимые словосочетания.
- 5) Метод и алгоритм обучения многослойной нейронной сети с переменной структурой, входным сигналом которой являются разряженные векторы большой размерности, обеспечивает.....
- 6) Аппаратно-программная реализация функций фильтрации незапрашиваемой рассылки электронной почты глобальной сети Интернет, определяет.....

Практическая ценность работы. Разработанные в диссертации модели текстовых сообщений, методы учета голосов пользователей и алгоритмы настройки нейронных сетей, а также аппаратно-программный комплекс, реализующий многоагентную АСФ НР, построенный на основе разработанных принципов позволяют:

- 1) повысить эффективность фильтрации потоков нежелательной информации;
- 2) упростить и ускорить цикл адаптации систем интеллектуальной фильтрации;
- 3) автоматизировать и сократить ручной труд при наладке системы;

4) предоставить пользователям гибкий механизм управления процессом фильтрации. Разработанные алгоритмы и программы могут быть использованы для дальнейшего развития и совершенствования различных интеллектуальных систем фильтрации и управления потоками текстовой информации.

Реализация результатов. Полученные в работе результаты: математические модели, методы и алгоритмы, а также разработанная АСФ НР внедрена и ежедневно используются, обеспечивая безопасность нескольких сотен (на момент публикации автореферата) пользователей корпоративных систем электронной почты в России и странах ближайшего зарубежья. Среди них:

- 1) распределенная почтовая служба Международного Института Экономики и Права (АСФ НР внедрена более чем в 10 филиалах).
- 2) почтовый портал ОАО “Концерн “РТИ-Системы”;
- 3) интернет-центр кафедры ИУ4 МГТУ им. Н.Э. Баумана.

Копии актов о внедрении прилагаются.

Внедрение результатов работы продолжается, они также могут быть использованы для создания широкого круга автоматизированных систем в различных прикладных областях: обеспечение безопасности, адаптация к профилю пользователя и т.д.

Апробация работы

Результаты работы были представлены на семинарах “Научоемкие технологии и интеллектуальные системы” в МГТУ им. Н.Э. Баумана.

Публикации

По материалам и основному содержанию работы имеется 12 публикаций, в том числе научно-технические статьи и доклады на конференциях.

Структура и объем работы

Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы. Общий объем диссертации 251 страниц, 62 рисунков, список использованных источников из 320 наименований.

КРАТКОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность решения задач борьбы с НР, сформулирована цель и задачи исследования, обоснована научная новизна и изложена структура диссертации, показано место АСФ НР рассматриваемого класса в современных корпоративных информационных системах.

В первой главе проведен анализ существующих АСФ НР, решающих задачу адаптивной фильтрации сообщений электронной почты, определены их особенности. Основные особенности заключаются в том, что данные АС должны: во-первых, функционировать на базе стандартной инфраструктуры электронной почты предприятия, во-вторых, обрабатывать большие объемы данных, передаваемых в интерактивном режиме, в-третьих, обеспечивать упрощенную адаптацию системы при обнаружении ошибок фильтрации, в-четвертых, реализовывать сложные многоитерационные алгоритмы обработки данных. Разработана их классификация в зависимости от архитектуры системы, способа ее интеграции с существующей инфраструктурой электронной почты, по режиму выявления ошибок фильтрации, по режиму настройки системы фильтров, по используемым методам фильтрации и математическому аппарату контекстного анализа (раздел 1.1). Данная классификация

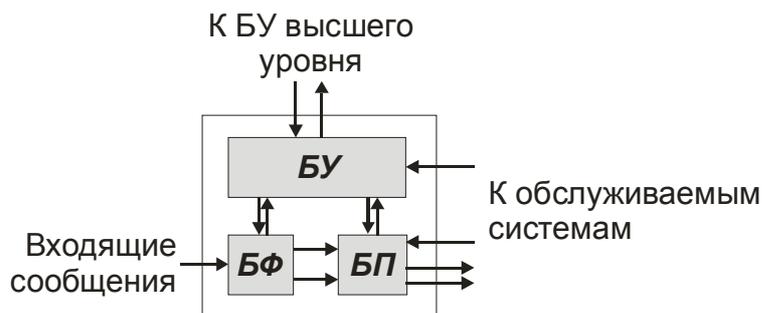


Рис. 1. Структурная схема агента АСФ НР

определяет требования к функциональным возможностям АСФ НР, изложенным в разделе 1.2.

Последний раздел главы посвящен описанию обобщенной методики построения многоагентной автоматизированной системы адаптивной фильтрации НР корпоративного уровня. Структура АСФ НР строится из агентов, имеющих идентичное структурно-функциональное построение (рис. 1). Блок фильтров (БФ) осуществляет разделение однородного потока входящих сообщений на два (поток НР и поток прочих сообщений). Блок предоставления (БП) реализуются на базе элементов стандартной инфраструктуры электронной почты предприятия и обеспечивает технические функции передачи потоков данных между элементами АСФ НР. Блок управления (БУ) обеспечивающим целостность и слаженность работы АСФ НР и управляет деятельностью прочих функциональных блоков в данном агенте.

Каскадное объединение агентов в сеть или иерархию позволяет построить АСФ НР практически любой конфигурации, отвечающей требованиям оперативного реагирования на ошибки фильтрации. На рис. 2 представлена типовая структура и потоки данных АСФ НР

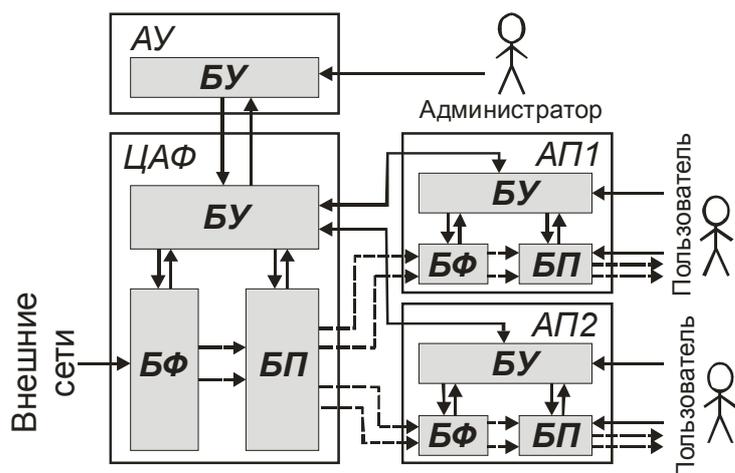


Рис. 2. Потоки данных в типовой схеме АСФ НР

корпоративного уровня, которая состоит из агентов трех видов:

- один центральный агент фильтрации (ЦАФ), обслуживающий всех пользователей корпоративной системы;
- множество агентов пользователей (АП), ориентированных на обслуживание одного пользователя;
- одного или нескольких агентов управления (АУ), представляющих собой вырожденный вид агента АСФ

НР (см. рис. 1) и служащих для реализации функций удаленного управления АСФ НР.

В разделе 1.3. описываются выполняемые отдельными агентами функции, принципы построения и взаимодействия отдельных функциональных блоков, а также функции, предоставляемые пользователям системы и административному персоналу.

В АСФ НР по функциональному признаку выделены три подсистемы: подсистема управления (ПУ), включающая БУ всех агентов АСФ НР; подсистема фильтрации (ПФ), состоящая из БФ; подсистема технического обслуживания (ПТО), состоящая из БП. Их целостное и скоординированное взаимодействие обеспечивается с помощью ведения глобального реестра сообщений (ГРС), который позволяет регистрировать и учитывать обработку отдельных сообщений в различных подсистемах АСФ НР.

Вторая глава посвящена исследованию и разработке математических моделей, методов и алгоритмов для решения задачи адаптивной фильтрации потоков текстовых данных, т.е. разработке математического обеспечения АСФ НР.

В разделе 2.1 рассматриваются математические методы контекстного учета поступивших в АСФ НР сообщений. Представлена хэш-функция, вычисляемая по тексту сообщения, позволяющая представить сообщение с помощью уникального 120-битного числа (идентификатора сообщения – ИС). ИС в частности позволяет выявить повторно поступающие в систему сообщения (полные текстовые дубликаты) даже при модификации его формальных параметров.

В разделе 2.2 рассматриваются математические методы коллективной фильтрации НР.

В разделе 2.2.1. представлены методы, позволяющие формировать общую выборку сообщений НР на основании поступающей от пользователей информации об ошибках фильтрации. Поскольку отдельные пользователи могут ошибочно относить некоторые

легитимные сообщения к классу НР, предусмотрена гибкая система, позволяющая учесть квалификацию отдельных пользователей.

В этом разделе определяется метод разделения общей выборки на две непересекающиеся выборки: обучающую (ОВ) и проверочную (ПВ). ОВ участвует в синтезе нейронной сети и формировании пространства признаков, ПВ используется для оценки качества обучения.

В разделе 2.2.2. представлены математические методы, позволяющие настраивать индивидуальные параметры фильтрации в зависимости от результатов голосования по отдельным сообщениям других пользователей системы. Данные методы позволяют изменять статус уже прошедших систему фильтров сообщений.

В разделе 2.3 представлен математический аппарат для решения задачи адаптивной фильтрации текстовых сообщений. В ее структуре выделены два основных элемента: кодирование и классификация (см. рис. 3). При кодировании сообщение представляется с помощью вектора в выбранном пространстве признаков. Классификация позволяет отнести сообщение к одному из классов (СНР или обычное) и осуществляется с помощью нейросети.

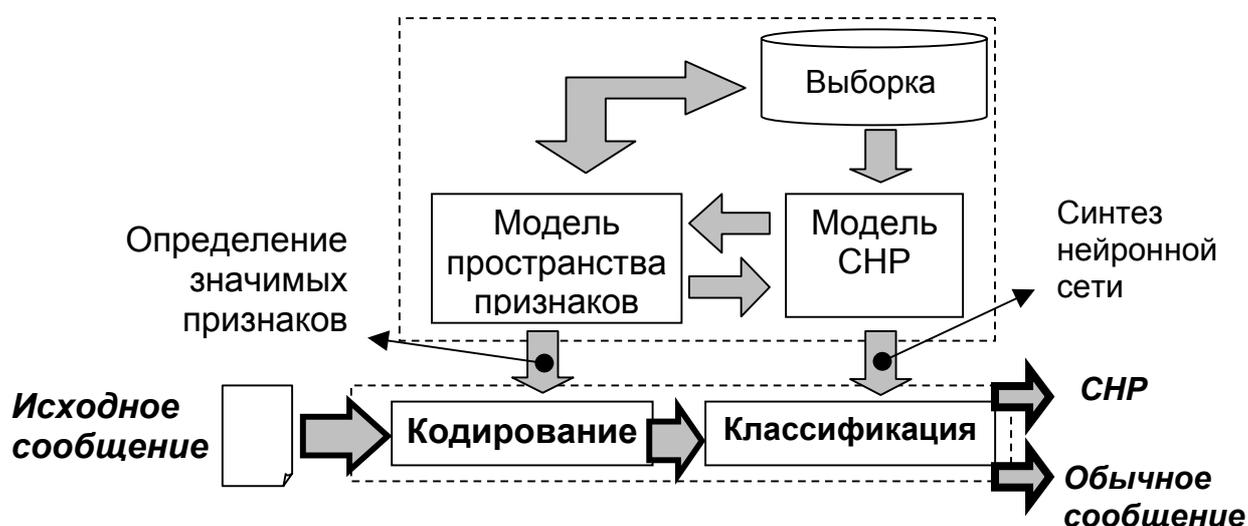


Рис. 2. Решение задачи адаптивной классификации

Синтез конкретного пространства признаков и нейросетевого классификатора осуществляется в процессе настройки системы на основании сформированной обучающей выборки.

В разделе 2.3.1 определяется метод формирования пространства признаков. В работе используется метод, при котором текстовый фрагмент представляется в виде точки n -мерного пространства: $X = (x_1, x_2, \dots, x_n)$, где каждая отдельная компонента соответствует либо слову общего словаря V , либо сочетанию слов словаря V . Под сочетанием здесь понимается множество слов (2-5 слов), встречающиеся в одном и том же словаре текстового фрагмента V_{ij} . Учет словосочетаний позволяет повысить точность классификации. Поскольку учет всех возможных словосочетаний невозможен, в работе осуществляется выбор наиболее значимых из них с использованием стохастического моделирования лингвистической структуры текстовых фрагментов, относящихся к одному классу в ОВ.

В разделе 2.3.1.1. описывается математический аппарат автоматического формирования словарей значимых слов. Эти слова далее используются в процедурах кодирования отдельных текстовых сообщений.

В разделе описывается процедура составления общего словаря системы V , в который попадают все слова из всех текстовых сообщений в ОВ. Перед занесением слова в словарь оно приводится к канонической форме (в соответствии с процедурой стемминга). Не выражающие смысла общие слова (это союзы, предлоги и т.п.) не вносятся в словарь, для чего предусмотрен их список. Занесение слова в словарь сопровождается подсчетом частоты его встреч в текстах ОВ.

В разделе описывается процедура сокращения общего числа различных слов в словарях. Для этого используются следующие (выбираемые в процессе настройки системы) показатели:

- Частота слов
- Information gain;
- Mutual Information;
- Статистика χ^2 ;
- Коэффициент корреляции.

Расчета показателей для всех слов общего словаря, в нем оставляют только заданное число слов, которые имеют наибольшие значения выбранного показателя (это число определяет производительность алгоритмов и является параметром, вводимым в систему в процессе настройки). На основании полученного таким образом общего словаря V , осуществляется составление частных словарей V_i , $i=1, \dots, K$ для каждого класса, а также V_{ij} для каждого текстового фрагмента.

Полученные таким образом словари определяют словарный состав, который будет учитываться в процессе классификации текстовых сообщений.

В разделе 2.3.1.2. вводится формальная модель лингвистической структуры категорий.

Эта модель определяется как вероятностная модель i -го класса в форме: $p(X_i | c_i)$, где c_i – символ i -ого класса, $X_i = (x_1, x_2, \dots, x_{|V_i|})$ – вектор, каждая компонента которого бинарна и соответствует слову из словаря V_i . Модель каждой категории строится независимо, поэтому для сокращения записи вводится обозначение: $P(X = x) = p(X_i | c_i)$. В качестве ОБ рассматривается TV , содержащая тексты общей ОБ, отнесенные к i -ому классу.

Для упрощения расчетов функция $P(X = x)$ рассматривается как функцию совместной плотности распределения n случайных величин X_i , каждая из которых определена на множестве $\{0,1\}$:

$$P(X = x) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n), \quad (2)$$

где $x_i = 0$, если i -ое слово присутствует в текстовом фрагменте X и $x_i = 1$ если наоборот.

В разделе 2.3.1.2.1. обосновывается возможность представления (2) в виде закона распределения Гиббса:

$$p(X = x) = \frac{1}{Z(\lambda_1, \lambda_2, \dots, \lambda_F)} e^{\sum_i \lambda_i f_i(x)} = \frac{1}{Z(\lambda)} e^{\lambda \cdot f(x)}, \quad (4)$$

где $f(x) = \{f_i(x), i=1, \dots, F\}$ – множество бинарных функций (признаки), λ_i – соответствующие им коэффициенты (параметры), $Z(\lambda) = \sum_x e^{\sum_i \lambda_i f_i(x)}$ – нормировочная постоянная. Функции признаки

являются либо простыми и соответствуют одному отдельному слову (т.е. обращаются в 1 тогда и только тогда, когда в векторе-аргументе соответствующая данному слову компонента обращается в 1), либо составными и соответствуют конъюнкции простых признаков.

В разделе 2.3.1.2.2. определяется задача формирования оптимального пространства признаков на основе вероятностной модели категории. Задача формирования пространства признаков решается с помощью выбора такого множества признаков $\{f_i\}$ и соответствующих им коэффициентов λ_i в модели (4), чтобы результирующее распределение $p(X=x)$ было максимально близко к эмпирической функции распределения $p'(X = x)$, которую получают из ОБ TV . Близость двух распределений в работе оценивается с помощью классических функционалов теории информации:

$$1. I(p': p) = \sum_x p'(x) \log \frac{p'(x)}{p(x)} \rightarrow \min, \quad (4.5)$$

и двойственного ему функционалу

$$2. I(p : u) = \sum_x p(X = x) \log \frac{p(X = x)}{u(x)} = \sum_x p(X = x) \log p(X = x) \rightarrow \min \quad (12)$$

при условии

$$\sum_x f_i(x) p(X = x | f) = \sum_x f_i(x) p'(X = x) = \theta_i, \quad i=1, \dots, F \quad (5)$$

где $\theta_i \in R, i=1, \dots, F$ - постоянные.

Далее в разделе описывается решение задачи с последовательного, многоитерационного алгоритма. Вначале определяется число признаков, которые необходимо добавить в модель (параметр алгоритма). На каждой итерации производится поиск очередного оптимального признака из числа кандидатов (структурный синтез). Далее найденный признак включается в модель, что достигается за счет оптимального выбора параметра λ_i , уточняющего модель с учетом нового признака (параметрический синтез). Алгоритм продолжается до тех пор, пока не будет добавлено требуемое количество признаков.

В разделе 2.3.1.2.3 описывается методы, использующиеся при структурном синтезе, т.е. в процессе выбора кандидата. Для этого применяется функционал (4.5) в форме:

$$I(f^*) = \inf_{\beta} I(p' : p(f^*, \beta)),$$

где f^* - оцениваемый кандидат, $p(f^*, \beta) = \frac{1}{Z'(f^*, \beta)} e^{\beta f^* + \sum_i \lambda_i f_i}$, $Z'(f^*, \beta) = \sum_x e^{\beta f^* + \sum_i \lambda_i f_i}$.

Поскольку данный функционал являются выпуклыми, то для их строго решения могут применяться классические методы теории оптимизации, такие как метод Ньютона, градиентного спуска и т.д.

В рассматриваемом бинарном случае:

$$f^*_{opt} = \arg \min_{f^* \in F_c} \left[\log \frac{1 - P^1(f^*)}{1 - E'(f^*)} + E'(f^*) \log \left(\frac{P^1(f^*)}{1 - P^1(f^*)} \frac{E'(f^*)}{1 - E'(f^*)} \right) \right],$$

где F_c - множество кандидатов, $P^1(f^*)$ - вероятность того, что признак f^* равен 1 относительно текущей модели $p(x)$, $E'(f^*)$ - среднее значение признака f^* на эмпирической выборке.

В разделе 2.3.1.2.4. описывается метод параметрического синтеза. При этом найденный оптимальный признак добавляется в модель, после чего осуществляется пересчет всех коэффициентов λ_i . Это достигается за счет условной минимизации функционала (12):

Для практической реализации алгоритма поиска минимума (12) используется ряд классических методов. В частности, метод улучшенного итеративного масштабирования, метод выборки из распределения Гиббса, а также метод Ньютона.

Приведенные подходы позволяют сформировать пространство признаков, включающее в качестве признаков как отдельные слова, так и словосочетания.

В разделе 2.3.2. описан применяющийся в работе метод синтеза многослойной нейронной сети с переменной структурой.

В разделе 2.3.2.1 приведена формальная постановка задачи оптимальной классификации с учетом специфики синтеза многослойной нейронной сети. Эта задача рассматривается как задача построения оптимальной гиперповерхности, разделяющей пространство векторов текстовых фрагментов X на ряд непересекающихся областей, соответствующих отдельным категориям. Если Y - пространство выходных сигналов, то оптимальное преобразование $X \rightarrow Y$ должно обеспечивать заданную вероятность отнесения входного вектора x к соответствующей ему категории. В данном разделе также определяется пространство указаний учителя E .

В разделе 2.3.2.2 вводятся основные соотношения, описывающие характеристики входного сигнала нейронной сети. В частности, определяются эмпирическая плотность распределения векторов k -ого класса: $f_k(x)$; априорные вероятности появления векторов из

k -го класса: p_k ; совместный закон распределения вероятностей входного сигнала x и указаний учителя ε : $f(x, \varepsilon)$; закон распределения указаний учителя $f_\varepsilon(\varepsilon)$; закон распределения входной совокупности векторов $f_x(x)$. Рассматриваются важные частные случаи: когда ни один из текстовых фрагментов не отнесен экспертами сразу к двум классам; случай конечной квалификации экспертов.

В разделе 2.3.2.3 вводятся и определяются различные критерии оптимизации, используемые в задаче классификации. В частности, рассматривается критерий максимума апостериорной вероятности $f(\varepsilon|x)$ и минимума средней функции риска R :

$$R = \sum_{i=1}^K \sum_{j=1}^K p_i l_{ij} \int \dots \int_{D_j} f_i(x) dx_1 \dots dx_N, \text{ где } l_{ij} - \text{коэффициенты матрицы потерь. Рассматриваются}$$

наиболее важные частные случаи данных критериев и характер оптимального решения задачи классификации, получаемой при их использовании.

В разделе 2.3.2.4 описан метод оптимального решения задачи классификации с применением математического аппарата нейронных сетей с переменной структурой. При использовании данного подхода разделяющая поверхность строится из отдельных линейных гиперповерхностей (соответствующих нейроподобным элементам - НЭ) в виде результирующей кусочно-линейной гиперповерхности. Синтез нейронной сети осуществляется на основании последовательного обучения слоев. Выбор параметров отдельных НЭ на каждом шаге алгоритма осуществляется в соответствии с минимизацией функционала среднего риска частного вида.

В разделе 2.3.2.5 описывается последовательный алгоритм синтеза первого слоя. Алгоритм строится в несколько этапов. На первом этапе производится построение части кусочно-линейной разделяющей гиперповерхности, отделяющей один выбранный класс от всех остальных $K-1$ классов. Затем рассматриваются только эти оставшиеся $K-1$ класса. Среди них снова выбирается класс и осуществляется построение очередной части кусочно-линейной гиперповерхности. Таким образом на каждом этапе решается задача отделения векторов двух выбранных классов.

Построение кусочно-линейной гиперповерхности, разделяющей два класса осуществляется с помощью последовательного алгоритма, который заключается в постепенном увеличении числа гиперповерхностей, до тех пор, пока не будет достигнуто заданное качество классификации или выполнено другое условие останова. Процесс обучения основывается на построении логического дерева. В геометрической интерпретации это выглядит следующим образом. Вначале исходное пространство признаков делится с помощью одного НЭ. Синтез НЭ осуществляется с помощью минимизации функционала среднего риска в пространстве коэффициентов данного НЭ. Затем на основании того же критерия полученные при делении области делятся еще раз и т.д. На рис. 3 представлена общая структурная схема алгоритма на каждом этапе: *I* – блок определения параметров НЭ; *II* – блок разделения входной обучающей последовательности; *III* (пунктир) – алгоритм синтеза многослойной нейронной сети на первом шаге, аналогично которому строятся блоки *III*; *VI* - блок управления.

Затем в разделе приводится структурная схема программы, реализующей алгоритм построения

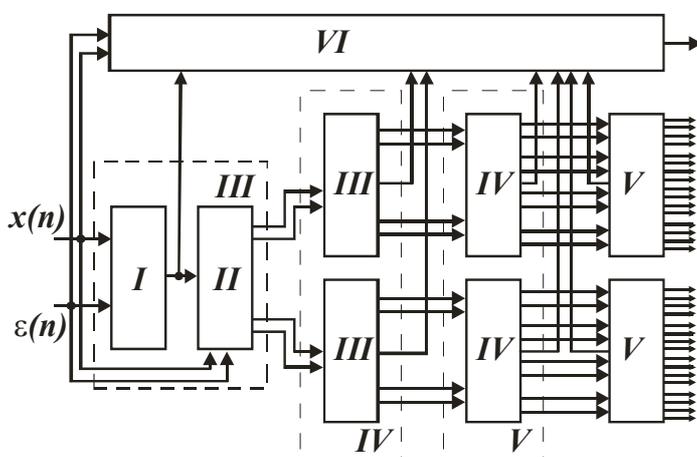


Рис. 3. Структурная схема алгоритма последовательного формирования кусочно-линейной разделяющей гиперповерхности

кусочно-линейной разделяющей поверхности.

Поскольку логическая функция, построенная на основе системы логических деревьев, оказывается определенной не на всех НЭ, далее в разделе описывается методика ее доопределения с целью формирования обучающей выборки для последующих слоев.

В разделе 2.3.2.6 определяется математический аппарат, использующийся при синтезе НЭ в алгоритме раздела 2.3.2.5 Синтез НЭ заключается в оптимальном выборе коэффициентов a_i в функции $y(x) = \text{sign}[\varphi(x)] = \text{sign}\left[\sum_{i=0}^N a_i x_i\right]$, реализующей данный НЭ. Выбор коэффициентов должен обеспечивать минимум функционала среднего риска.

Непосредственная минимизация этого функционала на поле коэффициентов a_i встречает существенные сложности. Поэтому, вместо решения прямой задачи в работе используется теоретически обоснованный переход к одной из эквивалентных задач, имеющих более естественное представление в пространстве коэффициентов НЭ. Для этого рассматриваются $x_a(n) = \varepsilon(n) - \varphi(x(n))$ - аналоговая ошибка НЭ; $x_o(n) = \varepsilon(n) - y(x(n))$ - дискретная ошибка.

Далее в работе рассматриваются моменты и модули моментов первого и второго порядка аналоговой и дискретной ошибки:

1. $\left|\overline{x_a(n)}^{m_n}\right| = \left|\overline{\varepsilon(n)}^{m_n} - \sum_{i=0}^N a_i \overline{x_i(n)}^{m_n}\right|;$
2. $\overline{x_a^2(n)}^{m_n} = 1 + \left[\sum_{i=0}^N a_i \overline{x_i(n)}^{m_n}\right]^2 - 2\varepsilon(n) \sum_{i=0}^N a_i \overline{x_i(n)}^{m_n};$
3. $|x_o(n)| = \left|\varepsilon(n) - \text{sign}\left[\sum_{i=0}^N a_i x_i(n)\right]\right|;$
4. $x_o^2(n) = \left(\varepsilon(n) - \text{sign}\left[\sum_{i=0}^N a_i x_i(n)\right]\right)^2,$

где использование соотношения $\overline{x(n)}^{m_n} = \frac{1}{m_n} \sum_{i=n-m_n}^n x(i)$ позволяет снизить уровень шумов в контуре настройки.

Далее в работе приводятся условия, при которых минимизация указанных выражений на поле коэффициентов, соответствует минимизации средней функции риска. В проектируемой АСАФ НР в процессе ее настройки и обучения предоставляется выбор одного из указанных критериев.

Минимизация приведенных выше критериев приводит к следующему рекуррентным соотношениям, позволяющим осуществить настройку НЭ:

1. $a(n+1) = a(n) + K^* \text{sign}\left[\overline{x_a(n)}^{m_n}\right] \overline{x(n)}^{m_n}$
2. $a(n+1) = a(n) + 2K^* \overline{x_a(n)}^{m_n} \overline{x(n)}^{m_n}$
3. $a(n+1) = a(n) + K^* \text{sign}[x_o(n)] \text{sign}(x_i(n))$
4. $a(n+1) = a(n) + 2K^* x_o(n) \text{sign}(x_i(n))$

где K^* - матрица, определяющая скорость настройки алгоритма.

В разделе 2.3.2.7 определяется метод синтеза последующих (второго и третьего) слоев. Данная задача определяется построением многомерной бинарной функции, которая должна реализовывать эталонную функцию $z(y) = (z_1(y), z_2(y), \dots, z_K(y))$, заданную на множестве бинарных векторов y , размерность которых совпадает с числом нейронов первого слоя H_1 . Поскольку осуществляется независимое построение отдельных компонент логической функций, реально решается задача построения одномерной бинарной функции, которая должна

совпадать с соответствующей $z_i(y)$. Для упрощения обозначений индекс опускается, т.е. вместо $z_i(y)$ записывается $z(y)$.

При синтезе $z(y)$ определяется можно ли ее реализовать с помощью одного НЭ. Если результат проверки оказывается положительным, то формируется нейронная сеть с единственным НЭ во втором слое. Если результат проверки отрицательный, то формирование $z(y)$ осуществляется с помощью дизъюнктивно-пороговой сети.

В разделе 2.3.2.8 приводится математический аппарат, использующийся для проверки реализуемости эталонной функции $z(y)$ на одном НЭ. Выходное значение НЭ определяется по

формуле $g(n) = \sum_{i=0}^{H_1} a_i y_i(n)$. Логическая функция $z(n)$ реализуема на одном нейроне, если можно

так подобрать коэффициенты a_i , что $\forall y(n)((z(n) < 0 \Rightarrow g(n) < 0) \wedge (z(n) \geq 0 \Rightarrow g(n) \geq 0))$. Это

эквивалентно тому, что у функционала $I(a) = \sum_{n=1}^M |g(n)| - \sum_{j=0}^{H_1} a_j b_j$, $b_j = \sum_{n=1}^M y_j(n) z(n)$, $j=0, \dots, H_1$

должен существовать минимум, в котором он обращается в ноль.

Таким образом, задача проверки условия реализуемости сведена к классической задаче минимизации функционала.

Оставшаяся часть раздела посвящена описанию метода минимизации данного функционала. Применяется два подхода. Во-первых, с использованием линейной аппроксимации, во-вторых, по таблицам пороговых функций.

Первый способ дает решение, при котором $a_i = b_i$, $i=1, \dots, H_1$, a_0 – варьируется для обеспечения минимума функционала. Второй применяется в случае, если число нейронов до 7 в соответствии с поиском решения по таблицам.

Если решение не может быть получено с использованием только одного НЭ, переходят к синтезу порогово-дизъюнктивной сети.

Раздел 2.3.2.9 посвящен вопросам построения порогово-дизъюнктивной двухслойной сети. Первый ее слой состоит из обычных НЭ, а второй из единственного НЭ вырожденного вида, реализующего схему *ИЛИ* (см. рис. 4).

В основе применяемого метода синтеза лежит использование понятия монотонной функции. Синтез двухкаскадной сети с элементом *ИЛИ* на выходе состоит в выделении ряда допустимых конфигураций, таких, что все единицы логической функции “покрываются”. Допустимая конфигурация есть подмножество единиц данной логической функции, которое может быть реализовано одним НЭ из входного каскада. В работе подробно описывается алгоритм, позволяющий осуществить синтез подобной сети.

Раздел 2.4 посвящен описанию математических методов, применяющихся для оценки качества построенной системы классификации. Данные методы используют обучающие и проверочные выборки с целью определения вероятности правильной классификации, которой достигла нейронная сеть в процессе обучения.

Третья глава посвящена созданию программно-аппаратного комплекса АСУ ИП, решающей задачу адаптивной фильтрации НР.

В разделе 3.1 определены требования к проектированию и реализации, учитывающие особенности АС данного класса. Данные требования явились определяющими при проектировании размещения типовых компонент АСФ НР (раздел 3.2).

Все функциональные блоки ЦАФ, а также БФ и частично БП всех АП реализуются на базе серверного аппаратно-программного комплекса, обслуживающего корпоративную систему ЭП в режиме 7x24. При этом функциональные блоки БП и БУ каждого АП реализуются на базе отдельного интеллектуального терминала в виде персональной ЭВМ конкретного пользователя.

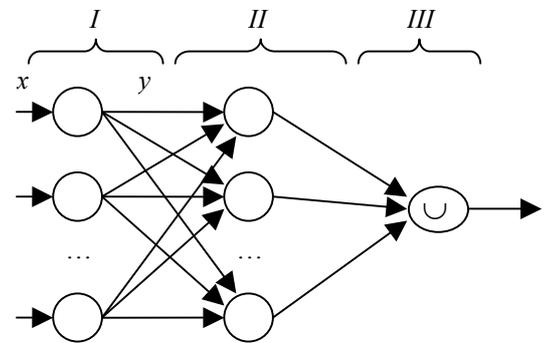


Рис. 4. Трехслойная сеть из НЭ

В серверной части БП АП организуется хранилище сообщений (ХС), в которое попадают сообщения после прохождения системы фильтров. В пользовательской части БП АП организуется доступ к сообщениям, хранящимся в ХС по запросу (протокол POP3 или IMAP4). При считывании сообщения из ХС предусмотрен вариант, когда сообщение остается в ХС, что позволяет многократно считывать одно и то же сообщение даже в том случае, когда ПЭВМ пользователя отключена.

В разделе 3.3 осуществляется выбор системного и стандартного прикладного программного обеспечения. Поскольку АСАФ НР являются развитием стандартных систем обработки сообщений, то для сокращения затрат на разработку системы, в качестве основы для АСАФ НР представляется целесообразным выбрать одну из существующих систем этого класса. В разделе обосновывается выбор в качестве платформы для построения почтового сервера системы Sendmail, а в качестве основы клиентской части – Microsoft Outlook.

Далее в разделе осуществляется выбор ОС. Выбор в качестве платформы системы Sendmail определяет выбор в качестве серверной ОС UNIX-подобную ОС. В работе анализируются другие полезные особенности ОС данного типа, в частности, указывается их гибкость и масштабируемость, что позволяет построить достаточную по производительности реализацию серверной части даже на базе недорогих аппаратных средств. В качестве клиентской ОС рекомендуется Windows XP SP1. Выбор в основном определяется высокими пользовательскими качествами данных систем, дружественным интерфейсом и широкой распространенности ОС данного типа, что гарантирует снижение затрат на внедрение и эксплуатацию системы.

Далее осуществляется выбор и обоснование средств разработки. Поскольку разрабатываемые программные средства применяются в гетерогенной среде, то в целях обеспечения программной совместимости и однородности модулей проекта, разработка должна отвечать требованиям кросс-платформенности. Другим требованием является масштабируемость языковых средств, которая с одной стороны должна позволять реализовывать дружественный интерфейс, с другой обеспечивать оптимальную реализацию высокопроизводительных многоитерационных алгоритмов. Наиболее данным требованиям отвечает язык С, который, во-первых, имеет реализации в большинстве существующих систем; во-вторых, средства разработки позволяют с одной стороны построить оптимальный код, с другой – обеспечивают средства быстрой разработки приложений (RAD) с пользовательским интерфейсом. Разработка основывается на стандартах ANSI, серверная часть следует стандартам Posix на базе RTL Glib и стандартной спецификации взаимодействия с системой Sendmail Milter, клиентская использует программный интерфейс приложений Win32, а также стандартную библиотеку языка С, совместимую с Posix. Взаимодействие компонентов системы обеспечивается с помощью спецификации программного интерфейса сокетов, совместимую с Posix.

Далее осуществляется выбор аппаратного обеспечения. Анализируются различные аппаратные платформы. В качестве рекомендуемой системы предложена система на базе процессора Intel Pentium 4, 512Мб оперативной памяти, объем жесткого диска до 80Гб в зависимости от числа пользователей.

В разделе 3.4. описываются типовые реализации системы на базе выбранных программно-аппаратных средств.

В разделе 3.5 подробно описываются модули клиентской и серверной части, их размещение и методы взаимодействия между новыми и стандартными модулями, информационные потоки между отдельными подсистемами. Также определяется программный интерфейс и протоколы взаимодействия.

В разделе 3.6 определяется структура баз данных, применяющихся в системе, которые систематизируются и классифицируются.

Раздел 3.7 посвящен краткому описанию методическому обеспечению АСУ ИП. Оно включает в себя справочную систему и документацию, определенную требованиями ГОСТ.

Четвертая глава посвящена описанию экспериментальных исследований, иллюстрирующих эффективность применения предложенных и разработанных моделей, методов и алгоритмов, решающих задачу адаптивной фильтрации текстовых сообщений. В разделе 4.1 описаны задачи, ставящиеся перед экспериментальными исследованиями, результаты которых описаны в последующих разделах данной главы.

В разделе 4.2. производится оценка качества классификации в зависимости от объема обучающей и проверочной выборки в различных сочетаниях. Качество классификации экспоненциально возрастает в зависимости от объема обучающей выборки (см. рис.). Такую же тенденцию демонстрирует проверочная выборка. Эксперименты показали, что наилучших результатов достигают систем, в которых общая выборка делится в пропорции: 2/3 – обучающая, 1/3 – проверочная.

Раздел 4.3 посвящен исследованию методов формирования пространства признаков. В экспериментах, описанных в этом разделе, рассматриваются два вида методов:

- метод предварительного сокращения пространства признаков (по критериям частоты слов, information gain, mutual information, хи-квадрат, коэффициента корреляции);
- метод учета значимых словосочетаний с использованием моделирования стохастической структуры категорий.

Установлено, что методы предварительного снижения размерности, основанные на частоте слов, information gain, хи-квадрат и коэффициенте корреляции позволяют сократить до 90% уникальных слов без потери качества или даже с некоторым его увеличением (до 2% в соответствии с мерой средней точности) – см. рис. 6. Все указанные методы можно рекомендовать в качестве альтернативы в процессе обучения. Однако, метод information gain рекомендуется в случае наиболее агрессивного сокращения пространства признаков; критерий частоты слов может использоваться для сокращения времени обучения; промежуточные результаты показали критерии хи-квадрат и коэффициент корреляции. Метод mutual information показал наихудшие результаты и не должен использоваться в системах данного класса.

Остальная часть раздела посвящена исследованию метода выбора значимых словосочетаний. Эксперименты показали, что качество классификации с использованием данного метода может быть повышено в пределах 2-5% в зависимости от числа учитываемых словосочетаний. Поскольку метод требует значительных объемов вычислений, его рекомендуется использовать только в случаях, когда в обучающей выборке происходят значительные изменения и не применять в штатных задачах дообучения.

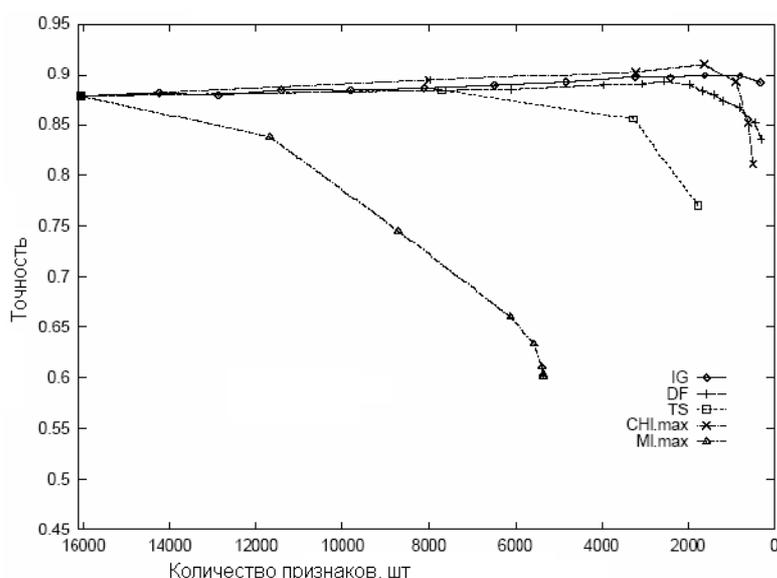


Рис. 6. Точность классификации при различных методах сокращения размерности пространства признаков

Раздел 4.4 посвящен оценке числа НЭ первого слоя, необходимых для реализации заданного качества классификации. Эксперименты показали, что увеличение числа НЭ первого слоя до 4-х на каждый класс приводит к росту вероятности правильной классификации в пределах в среднем от 4% (2 НЭ 1-го слоя) до 5,8% (4 НЭ первого слоя). Дальнейшее увеличение числа нейронов в большинстве экспериментов показало снижение вероятности правильной классификации на проверочной выборке, что объясняется эффектом переобучения (overfitting) нейронной сети.

В разделе 4.5 изучена работа

различных методов, предлагаемых системой для обучения НЭ первого слоя. В экспериментах, описанных в этом разделе, рассматриваются два фактора, влияющих на качество классификации:

- выбор критерия вторичной оптимизации (первый и второй моменты дискретной и аналоговой ошибки на выходе НЭ);
- выбора шага процедуры настройки коэффициентов НЭ.

Сравнение и оценка рассматриваемых критериев вторичной оптимизации показала близкие результаты для каждого из них. В частности, эксперименты с нейронными сетями, с 4-мя НЭ на каждый класс в первом слое показал результаты, представленные на рис.

Эксперименты с выбором шага проводились для следующих режимов:

- постоянный шаг;
- монотонно убывающий шаг;
- адаптивный шаг.

Наилучшие результаты показала процедура с адаптивным шагом, позволяющая повысить качество классификации в пределах до 6%.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ

Разработаны принципы и методы построения многопользовательских автоматических многоагентных систем фильтрации потоков текстовых сообщений, функционирующих на базе адаптивных алгоритмов, а также их программная реализация.

- 1) Проведен анализ и систематизация возможностей существующих корпоративных АСФ НР и методов их интеграции в системы электронной почты.
- 2) Выработаны требования к функциональному составу, алгоритмическому и программному обеспечению, выполнение которых необходимо для построения эффективных в современных условиях АСФ НР.
- 3) На базе стандартной инфраструктуры систем электронной почты разработана и исследована многоагентная архитектура АСФ НР, позволяющая отдельным пользователям участвовать в процессе фильтрации НР: воздействовать на алгоритмы настройки фильтров и осуществлять выбор степени жесткости фильтрации в индивидуальном порядке.
- 4) Разработан комплекс средств, позволяющих производить сквозной учет поступающих в корпоративную систему сообщений.
- 5) Разработаны математические методы, позволяющие производить формирование обучающей выборки с учетом значимости голосов отдельных пользователей.
- 6) Разработаны математические методы выбора пространства признаков в задаче анализа содержания сообщений, что позволило повысить точность анализа за счет учета значимых словосочетаний.
- 7) Разработан метод синтеза нейронной сети с переменной структурой, входным сигналом которой являются разряженные векторы большой размерности (десятки тысяч).
- 8) Проведены исследования и выбор методов программной и аппаратной реализации средств фильтрации НР.
- 9) Разработан аппаратно-программного комплекс фильтрации НР на основе разработанных принципов;
- 10) Проведены экспериментальные исследования, наглядно демонстрирующие эффективность применения разработанных принципов, моделей, методов и алгоритмов в задаче фильтрации НР в информационной сети Интернет. Средняя вероятность правильной фильтрации сообщений НР по профилям “жесткая”, “умеренная” и “мягкая фильтрация” соответственно составила: 68,2%, 79,6%, 82,0%. Вероятность ложной фильтрации пользовательских сообщений соответственно: 0%, 0,01%, 1,4%. Для сравнения, современные коммерческие системы в режиме с уровнем ложных сработок порядка 2-3% достигают пикового показателя вероятности правильной фильтрации 70-75%.

Основные положения диссертации изложены в следующих работах:

1. Цыганов И.Г. Нейросетевые методы автоматизированного анализа информационных потоков в масштабе реального времени Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2002 г., с. 19-24
2. Цыганов И.Г. Оценка применимости нейросетевых парадигм при решении задачи сквозного семантического анализа текстовых сообщений V Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2003 г., Часть I, с. 66-77
3. Цыганов И.Г. Генерирование тестовой выборки большой размерности и модальности в задаче самообучения нейронной сети V Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2003 г., Часть I, с. 78-86
4. Цыганов И.Г. Neural Network Hardware-Software Complex for Raw Message Stream Clusterization in Real Time Mode V Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2003 г., Часть II, с. 120-125
5. Цыганов И.Г. О динамике настройки нейронной сети при решении задачи кластеризации входного сигнала большой размерности и модальности V Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2003 г., Часть II, с. 126-131
6. Цыганов И.Г. Применение нейросетевых методов для фильтрации SPAM сообщений Международная молодежная научная конференция “Информатика и системы управления в XXI веке”. Сборник научных трудов. М., 2003 г., с. 26-33
7. Руденко М.И., Цыганов И.Г. Метрики текстов в автоматизированных системах обработки информации, VI Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2004 г., с.
8. Цыганов И.Г. Метод аналитической настройки нейронных сетей в задаче контекстной классификации текстов, VI Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2004 г., с.
9. Цыганов И.Г. Решение задачи автоматизированной контекстной классификации с помощью стохастического моделирования лингвистической структуры категорий, VI Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2004 г., с.
10. Цыганов И.Г. Формирование пространства признаков в задаче категоризации интерактивных потоков текстовой информации, VI Международная молодежная научно-техническая конференция “Научно-технические технологии и интеллектуальные системы”. Сборник научных трудов – М., 2004 г., с.
11. Цыганов И.Г., Власов А.И. Адаптивная фильтрация информационных потоков в корпоративных системах на основе механизма голосования пользователей // Информационные технологии, Сентябрь 2004.
12. Цыганов И.Г., Власов А.И. Архитектура корпоративной многоагентной автоматизированной системы фильтрации информационных потоков // Информационные технологии, Январь 2005.