Таким образом САЗ ПП реализованные в соответствии с схемой (рис. 5) могут использоваться для активной маскировки всех основных типов опасных сигналов, характерных для современных корпоративных сетей, с учетом ограничения по тактовой частоте, накладываемых формулой (2).

По своим техническим характеристикам системы активной защиты могут быть использованы для защиты практически любых технических средств корпоративных сетей.

Однако сложность САЗ возрастает пропорционально количеству опасных сигналов, подлежащих маскировке, и количеству узлов и блоков, имеющих высокие уровни побочных излучений.

Преимуществом методов активной защиты по сравнению с методами пассивной защиты является то, что для каждого технического средства корпоративной сети, требуемый уровень защиты может обеспечиваться индивидуально. При общем же экранировании существует проблема сочетания в одном комплексе устройств с разными уровнями защиты.

Список литературы

1. *Защита* от радиопомех/Под ред. М.В. Максимова.- М.: Сов.радио, 1967. -496 с.

2. *Нормы* эффективности защиты АСУ и ЭВМ от утечки информации за счет ПЭМИН. -М.: МОП, 1977. -35 с.

3. *Специальные* требования по защите объектов ЭВТ второй категории от утечки информации за счет ПЭМИН. – М.: Гостехкомиссия СССР, 1979. - 44 с.

4. *Специальные* требования по защите объектов ЭВТ третьей категории от утечки информации за счет ПЭМИН. – М.: Гостехкомиссия СССР, 1979. - 43 с.

5. *Хмелевский И.В.* Анализ эффективности использования аддитивных помех для маскировки передачи речи методом дельта-модуляции.- Диссертация кандидата технических наук. Свердловск. -1989. -182 с.

6. *Гуткин Л.С.* Теория оптимальных методов радиоприема при флюктуационных помехах: 2-е изд.- М.: Сов радио,1972. -448 с.

7. *Левин Б.Р.* Теоретические основы статистической радиотехники: в 3-х томах, т.2.- М.: Сов.радио, 1975.-552 с.

8. *Новиков А.А.* Нормированные величины сигнал-помеха для сигналов с частотной и фазовой модуляцией, применяемых в системах телеобработки информации // Вопросы специальной радиоэлектроники. Серия "Электронная вычислительная техника", 1982, вып.3,- с.84-88.

9. *Вакман Д.Е.* Асимптотические методы в линейной радиотехнике. -М.: Сов.Радио, 1962.-373 с.

10. *Вайнштейн Т.Г.* Теория обработки сигналов автоматического управления в радиоэлектронных системах.- Л.: МО, 1992.-245 с.

11. *Втаонкин В.М., Каркаускас Ш.М.* Генератор импульсного случайного потока //Труды Рязанского политехнического института, вып.64.Рязань, 1975. - с.17-25.

12. *Левин Б.Р.* Теоретические основы статистической радиотехники: в 3-х томах, т.1.- М.: Сов. Радио, 1974.-392 с.

13. *Тихонов В.И.* Выбросы случайных процессов.-М.: Наука, 1970.- 375 с.

14. *Шеннон К.Э.* Работы по теории информации и кибернетике /Пер. с англ. -М.: Иностранная литература, 1963.-829 с.

15. *Положення* про технічний захист інформації в Україні. Затверджено Указом Президента України від 27.09.99. № 1229.

Поступила 29.08.2006

УДК 681.3.06

Терейковский И.А.

ПРИМЕНЕНИЕ СЕМАНТИЧЕСКОГО АНАЛИЗА СОДЕРЖИМОГО ЭЛЕКТРОННЫХ ПИСЕМ В СИСТЕМАХ РАСПОЗНАВЕНИЯ СПАМА

Введение

В общем случае, под термином спам, понимают массово рассылаемые электронные письма, содержащие информацию, которая не интересует большинство получателей или вводит их в заблуждение. Как правило, такие письма анонимны, предназначены для достаточно широкого круга лиц, при этом большинство пользователей не подписывались на получение данной почтовой рассылки. По своей сути большая часть спама является навязчивой и несанкционированной рекламой. Уверенность руководителей многих компаний в том, что такой метод рекламы достаточно эффективен, является основной предпосылкой существования спама. Предполагается, что большинство получателей охотно откликаются на содержащиеся в спаме предложения товаров и услуг. Кроме этого, спам это один из самых дешевых способ рекламы. Стоимость контакта с клиентом при массовой рассылке рекламных писем гораздо ниже, чем при рекламировании любым другим способом [1,2]. Поэтому рассылка спама это высокодоходный бизнес, подкрепленный устоявшимся рынком и стабильным спросом. Следствием этого является то, что в русскоязычной зоне Интернет объем спама составляет около 80% от общего объема всей электронной почты [1,3].

С точки зрения конечных пользователей, основными отрицательными моментами существования спама являются уменьшение эффективности обработки содержимого электронных писем и увеличение трафика при использовании электронной почты. Если не учитывать психологический аспект получения письма от неизвестного источника, то снижение эффективности обработки электронной корреспонденции происходит за счет временных потерь на обработку спама. По данным [1,3] в крупных компаниях сотрудники, которые работают с электронной почтой, тратят на спам около трех процентов своего рабочего времени. Негативные последствия спама привели к тому, что в некоторых странах борьба со спамом ведется уже и на законодательном уровне. Этим объясняется актуальность общей проблемы данной статьи – исследования методов и средств защиты от спама. Отметим, что основной проблемой защиты от спама является его распознавание в общем потоке получаемых электронных писем.

Анализ современных методов распознавания спама

Метод "черного", "белого" и "серого" списков. Базой метода является анализ обратного IP-адреса отправителя письма. Все письма, отправленные с IP-адресов, занесенных в "черный список", уничтожаются еще на почтовом сервере, так и не достигая конечного пользователя. Адрес вноситься в "черный список" на основании того, что письмо пришедшие с этого адреса является спамом. С адресатами из "белого списка" разрешен обмен почтовыми сообщениями. В случае, когда IP-адрес письма не присутствует ни в "черном" ни в "белом" списке, то отправителю автоматически высылается запрос на авторизацию, а IP-адрес заноситься во временный "серый" список. Если по истечении определенного срока подтверждение "благонадежности" от неизвестного адресанта не поступает, то его адрес вносится в "черный список", а сообщения удаляются. Основной недостаток данного метода заключается в том, что IP-адрес не обязательно является указателем источника спама. Например, спам может прийти с динамического IP-адреса, или рассылка спама может производится без санкции владельца IP-адреса. Таким образом, с высокой вероятностью в "черный" список могут попасть адреса ни в чем не повинных пользователей. Кроме этого использование "серого" списка оправдано только в том случае переписки с узким кругом лиц. Если же пользователю приходится работать с электронной

почтой достаточно много и часто поступают письма от неизвестных людей, то ведение "серого" списка потребует достаточно больших затрат на периодическую реконфигурацию.

Метод писем-подтверждений. В некотором смысле является модификацией метода списков. При использовании метода подтверждений в ответ на получение письма IP-адрес, которого не внесен в "белый" список высылается запрос с просьбой подтверждения факта отправки. В случае подтверждения IP-адрес заносится в "белый список", а исходное письмо доставляется получателю. Метод базируется на том, что поскольку спам-рассылки происходят автоматически, по многим миллионам адресов, а адрес отправителя - в большинстве случаев – поддельный, то подтверждения от настоящего спамера получить не удастся. Однако применение данного метода резко снижает оперативность доставки писем, во многих случаях спам отправляется с реальных IP-адресов, а современное программное обеспечение спамеров может генерировать подтверждение отправки писем.

Метод распознавания спама по ключевым словам (словосочетаниям), которые определяются пользователем в виде некоторых правил. Данный метод не получил широкого распространения в силу сложности и трудоемкости формирования указанных правил.

Метод байесовской фильтрации. Каждому встречающемуся в электронной переписке слову (или HTML-тэгу) присваивается два значения: вероятность его наличия в спаме (z) и вероятность его присутствия в письмах, разрешенных для прохождения (1-z). Каждому новому письму с помощью формулы Байеса выставляется оценка (Z):

$$Z = A/(A+B), \tag{1}$$

где

$$A = z_1 \times z_2 \times \ldots \times z_i \times \ldots \times z_n, \tag{2}$$

$$B = (1-z_1)\times(1-z_2)\times\ldots\times(1-z_i)\times\ldots\times(1-z_n), \tag{3}$$

z_i - спам-оценка каждого слова, входящего в письмо.

Если полученная оценка меньше некоторого заранее определенного порогового значения, то письмо трактуется как спам.

Очевидно, что эффективность данного метода во многом зависит от правильности расчета спам-оценок слов входящих в письмо. Для этого необходимо произвести статистический анализ как спама, так и обычных писем получаемых каждым пользователем. Необходимость индивидуального анализа объясняется:

- Пользователи могут иметь различные интересы. Для одних пользователей письмо является спамом, для других оно представляет интерес.
- Разные пользователи используют при переписке различную лексику.

Таким образом, метод байесовской фильтрации предполагает некоторое запаздывание, связанное с накоплением каждым пользователем достаточного объема статистического материала (архива писем). Еще одним недостатком метода является пропуск спама, если в письме относительно мало слов с высокой спам-оценкой. Отметим, что это обстоятельство используется спамерами как для обхода, так и для компрометации фильтров. Например, бессмысленное письмо, состоящее из набора нейтральных слов, не будет распознано как спам.

В большинстве современных антиспамовых систем реализованы комплексные методы защиты, которые по заверениям их разработчиков могут фильтровать до 98% спама. Однако время реакции на новый вид спам-писем крупнейших почтовых служб Интернета составляет 20-30 мин [1,3]. Отметим, что эти почтовые службы защищены наиболее современными средствами защиты. При этом, крупные рассылки многих миллионов спам-писем осуществляются в течении 1-2 часов. Поэтому с большой вероятностью в почтовые службы многих пользователей проведут не верную классификацию спама.

Проведенный анализ позволяет сформулировать вывод о том, что существующие системы распознавания спама не могут адекватно реагировать на современные методы составления и распространения спама. В то же время, даже не квалифицированный пользователь легко распознает спам на основании сопоставления своих интересов и смысла

письма. По этой причине целесообразно распознавать спам по аналогии с тем, как это делает человек, т.е. на основании анализа содержания письма.

Цель статьи

Усовершенствование методик распознавания спама на основании анализа содержания электронных писем.

Концепция фильтрации электронных писем

Сам факт существования достаточно дорогостоящих массовых рассылок электронных писем рекламного характера свидетельствует о том, что для многих пользователей Интернета спам представляет большой интерес. Очевидно, что этот интерес обусловлен содержанием спам-писем. В то же время пользователи, которые не интересуются предложенной тематикой, относятся к спаму отрицательно. По этим причинам основным критерием фильтрации электронных писем может быть соответствие содержания электронного письма и интересов пользователей:

$$\begin{cases} \forall P,T \in \{I\} \to C \\ \forall P,T \notin \{I\} \to S' \end{cases} \quad (4)$$

где P– электронное письмо, T– тематика электронного письма, $\{I\}$ – множество (область) интересов пользователей, C – целевое письмо, S – спам.

Исходя из возможностей потенциальных эксплуатантов системы защиты, формирование области интересов пользователей необходимо реализовать с помощью одного или нескольких фрагментов текста на естественном языке. В качестве указанных фрагментов могут использоваться специальным образом обработанные целевые письма, а также непосредственно введенный текст. Возможной проблемой реализации зависимости (4) является определение эксплуатантами системы защиты, всей области интересов пользователей электронной почты. На практике может оказаться, что даже конечному пользователю четко определить границы этой области достаточно трудно. При этом границы области интересов могут изменяться во времени. Поэтому многие потенциально интересные письма могут быть расценены как спам. Для решения данной проблемы разделим все электронные письма на три группы: целевые письма, спам и нейтральные письма. В группу подозрительных будут попадать те письма, тематика которых не принадлежит ни множеству интересов пользователей, ни множеству тематик спама. Учитывая предложенную классификацию, модифицируем критерий фильтрации (4):

$$\begin{cases} \forall P,T \in \{I\} \to C \\ \forall P,T \notin \{I\} \wedge T \notin \{Q\} \to F, \\ \forall P,T \notin \{I\} \wedge T \in \{Q\} \to S \end{cases} \quad (5)$$

где F – нейтральное письмо, $\{Q\}$ – множество тем спама.

Тематика спама

Практический опыт, а также результаты [1,2,3] показывают, что за несколько последних лет спам представляет собой в основном текстовые письма, которые иногда имеют графические файлы-вложения. При этом основными тематическими направлениями спама являются [1,2,3]:

– Реклама потребительских товаров (R_t). Рекламируется реальный товар, и указываются источники (ссылки на сайт или номер телефона) более подробной информации. Интересной особенностью этого направления спама является доминирование в определенные интервалы времени рекламы конкретного вида товара. Например, в русскоязычной зоне Интернет в период 2004 года реклама лекарственных препаратов составила около 17% количества спама.

– Реклама товаров и услуг "для взрослых" (R_p).

– Реклама программного обеспечения и компьютеров (R_k).

– Реклама туристических компаний, предлагающих различные виды отдыха и путешествий (R_o).

– Приглашения на семинары и тренинги (R_{st}).

– Услуги по электронной рекламе (R_{er}).

– Платные звонки. Рекламируется товар и/или услуга и указывается номер телефона, звонки на который являются платными (R_z).

– Раскрутка сайта. Письмо содержит информацию с целью завлечь пользователей посетить определенный сайт (R_w).

– Финансовый спам. К этому виду спама относятся письма с рекламой различного вида денежных пирамид, предложения сделать определенную инвестицию или реклама покупки акций (R_f).

– Сбор информации. Получателю предлагают заполнить анкету и отослать данные по указанному адресу (R_i).

– Политические или PR-акции. Этот вид спама характерен в периоды обострения политической обстановки (R_{pr}).

– Засылка троянов. При открытии письма активизируется программа типа троянский конь, которая выполняет некоторые несанкционированные действия, например, собирает и отсылает злоумышленнику необходимую информацию с компьютера (W_t).

– Фишинг. Это распространение поддельных сообщений от имени банков/финансовых компаний. Целью такого сообщения является несанкционированный сбор идентификационных данных (паролей, пин-кодов, логинов) пользователей. Обычно такой спам вынуждает пользователя ввести свои идентификационные данные, например пароль для доступа к банковскому счету на ложном сайте банка. Полученные идентификационные данные спамер может использовать как для доступа к счету, так и для оплаты покупок в интернет-магазинах (W_f).

– Тестовые рассылки. Чаще всего представляют собой пустые письма (T_p), письма с несколькими словами (T_s) или с бессмысленным набором символов (T_b). Такие рассылки преследуют сразу несколько целей. С одной стороны, это обычное тестирование нового или модифицированного спамерского программного обеспечения. С другой стороны, письма таких рассылок достаточно часто проходят антиспам-фильтры (не содержат спамерского контента), вызывая у пользователей недоверие к защите от спама. Еще одно негативное свойство тестовых рассылок связано с созданием ими больших дополнительных нагрузок на каналы связи. Это может выражаться в существенном снижении скорости обмена электронной корреспонденцией на время прохождения рассылки.

Модифицируем (5), с учетом распространенных тем спама:

$$\begin{cases} \forall P,T \in \{I\} \rightarrow C \\ \forall P,T \notin \{I\} \wedge T \notin \{R_t,R_p,R_k,R_o,R_{st},R_{er},R_z,R_w,R_f,R_i,R_{pr},W_t,W_f,T_p,T_s,T_b,N\} \rightarrow F \\ \forall P,T \notin \{I\} \wedge T \in \{R_t,R_p,R_k,R_o,R_{st},R_{er},R_z,R_w,R_f,R_i,R_{pr},W_t,W_f,T_p,T_s,T_b,N\} \rightarrow S' \\ \{R_t,R_p,R_k,R_o,R_{st},R_{er},R_z,R_w,R_f,R_i,R_{pr},W_t,W_f,T_p,T_s,T_b,N\} \in Q \end{cases} \quad (6)$$

где N – спам-письма с тематикой, не принадлежащей ни к одной из выше перечисленных распространенных тем спама.

Задача определения соответствия смысла электронных писем с интересами пользователей или с тематикой спама

На наш взгляд указанная задача может быть отнесена к классу задач общения человека с вычислительной системой на естественном языке. В настоящее время, несмотря на значительные успехи эта проблема далека от решения. Поэтому поиск решения следует

ограничить, соизмерив с существующими возможностями методик понимания текста и потребностями системы защиты от спама. Следует учитывать, что система распознавания не обязательно должна понять смысл текста электронного письма, интересы пользователя и тематики спама. Задача состоит в том, что бы сравнить формальное описание смысла указанных текстов и отнести электронное письмо к одному из заранее известных классов.

В настоящее время, перспективным методом извлечения смысла текста является реферирование [4,5]. Поэтому, и при классификации электронных писем, возможно, использовать рефераты эти писем. С учетом этого предположения, сформирован алгоритм классификации электронных писем, показанный на рис.1.

Отметим, что в алгоритме не учтены технические моменты, связанные с открытием электронного письма, определением кодировки символов и т.д. Кроме этого, предполагается, что определить тематику письма возможно на основании анализа его текста. Достоинствами применения рефератов являются:

- Составление рефератов текста достаточно хорошо отработано как на теоретическом, так и на практическом уровне. Во многих случаях рефераты удовлетворительно отображают смысл представляемых текстов.
- Возможность автоматического распознавания и блокировки бессмысленных писем, которые практически не распознаются большинством современных систем защиты от спама.
- Формальное представление реферата в большинстве случаев гораздо короче, а значит и требует гораздо меньшего объема ресурсов (памяти) для хранения, чем формальное описание исходного текста.
- Сопоставление относительно коротких рефератов в значительной мере уменьшат трудности, связанные с многообразием языковых форм.

Задача сопоставления рефератов

Основная трудность при сопоставлении рефератов заключается в том, что практически одинаковый смысл может быть выражен с помощью разного количества слов, довольно большого количества различных языковых конструкций, словосочетаний, слов синонимов. Вопрос несколько упрощается из-за того, что рефераты могут быть созданы по одинаковым правилам, учитывающим необходимость уменьшения применяемых языковых конструкций. При этом большинство современных методик формирования рефератов базируются на использовании семантических сетей. В общем случае семантическая сеть представляет знания в виде графа, узлы которого соответствуют фактам, а дуги – отношениям или ассоциациям между понятиями. Достоинством семантических сетей является возможность определения связей между понятиями и специфических правил вывода, определяемых механизмом наследования.

Под семантической сетью текста понимают множество связанных между собой значимых понятий (слов и словосочетаний), выделенных из состава текста. Каждый элемент семантической сети (понятие) характеризуется своим весом и набором связей с другими элементами – контекстным узлом. Вес элемента определяет относительную смысловую значимость выраженной им темы по сравнению с значимостью других элементов. Вес связи между парой элементов характеризует относительную смысловую связность, соответствующей первому элементу, с темой соответствующей второму. Каждый вес элемента и вес связи характеризуется числовым значением в заранее определенном диапазоне.

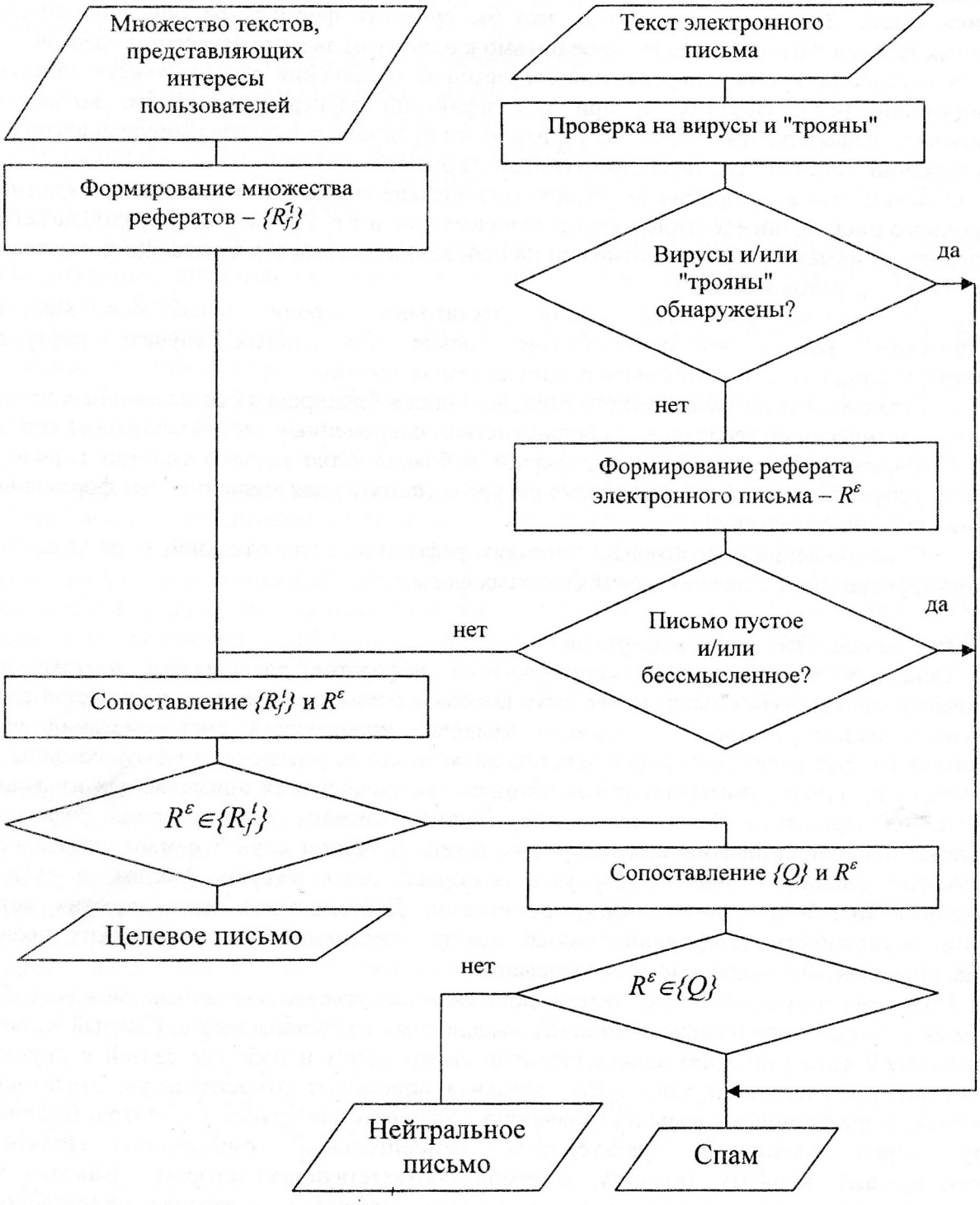

Рис. 1. Укрупненный алгоритм классификации электронных писем

Применение семантических сетей позволяет абстрагироваться от малоинформативных элементов формально-синтакисической структуры текста (порядка слов, залога и т.п.) и представляет его пропозициональную структуру в терминах описываемых ситуаций (предикатов) и их участников (аргументов) в определенных семантических ролях [4, 5, 6]. Однако, в задаче распознавания спама, полное представление смысла текста в форме семантической сети является избыточным и непродуктивным. Такое представление имеет большой объем (превышающий объем документа), а его обработка требует развитых нетривиальных средств для поиска и сравнения структур на графах, что в свою очередь требует использования значительных вычислительных ресурсов. Возможным выходом из этой ситуации является представление смыслового портрета в виде перечня элементарных смыслов - атрибутов, с оценками их информативности для характеристики текста. Традиционно в силу простоты реализации для этой цели используются частотные списки слов, которые употребляются в тексте. Однако наиболее информативные элементы смысла, описывающие отношения, возникают только на уровне синтагм, выделение которых требует применения алгоритмов синтаксического анализа, описанных, например, в [5]. Будучи дополнен правилами для генерации канонической формы синтагм, синтаксический анализ-синтез позволит описать каждый смысловой атрибут текста в виде строки, инвариантной к его грамматическому выражению в различных фразах.

Возможность составления качественных рефератов подтверждаются распространенностью соответствующих программных продуктов, например TextAnalyst, компании "Микро Системы" и Inxight Summarizer, созданной в Исследовательском центре Ксерокса. К задачам, решаемым подобными программными продуктами, относятся:

- Построение словарей базовых понятий предметной области (терминологических словосочетаний и слов) на базе множества текстов.
- Построение баз знаний на базе текстов предметной области с оценкой относительной значимости понятий и их смысловых связей.
- Построение тематической структуры текста. Тематическая структура описывает содержание анализируемых текстов в виде иерархии связанных тем и подтем, раскрывающих содержание тем.
- Построение списка наиболее значимых предложений (реферата) исходного текста. При этом каждое предложение реферата может характеризоваться степенью значимости.
- Рубрикация (классификация) текстов по заданным темам.
- Смысловой поиск интересующей информации. Функция смыслового поиска позволяет получить ответ на запрос, сформированный в виде фразы естественного языка, словосочетаний или просто набора ключевых слов. При этом извлекаемая в ответ на запрос информация может иметь другую грамматическую форму или вообще не упоминаться явно в тексте запроса, однако иметь смысловую связь с текстом запроса.

Отметим, что решение этих же задач необходимо и при классификации электронных писем. Автором были проведены эксперименты по составлению рефератов спам-писем по тематикам: приглашения на семинары и реклама потребительских товаров. В качестве инструмента реферирования был использован TextAnalyst.- Результаты экспериментов показали удовлетворительное качество, как составления тематической структуры анализируемых писем, так и формирования списка наиболее значимых предложений (реферата). При этом объем файла реферата составил около 5-10% от объема файла анализируемого текста. Кроме этого, сделана попытка составления реферата пустого файла и бессмысленных текстов. Система TextAnalyst выдала сообщение о том, что данные тексты не подлежат реферирования, так как являются неправильными. В интерпретации системы, письма с таким содержимым были бы классифицированы как спам что, безусловно, является положительным моментом.

В то же время, возможности смыслового поиска и рубрикации текста не отвечают потребностям системы защиты от спама. Так смысловой поиск слова строение в письме,

посвященному рекламе жилья закончился безрезультатно. Результаты [4, 5, 6] показывают, что качественно решить вопросы рубрикации и смыслового поиска возможно за счет сопоставления рефератов и/или тематической структуры текстов с использованием грамматических словарей. Для проведения такого сопоставления, возможно, использовать рекуррентные семантические нейронные сети или вероятностные нейронные сети [6, 7]. Сравнение указанных типов нейронных сетей показывает, что рекуррентные семантические нейронные сети обладают большей производительностью и мощностью в задачах классификации и кластеризации образцов текстов. Однако их реализация в системах распознавания спама требует проведения дополнительных исследований. Кроме этого практическая реализация рекуррентных семантических сетей не всегда возможна по причине использования значительных вычислительных ресурсов. Поэтому в данной статье ставиться акцент на применении вероятностных нейронных сетей.

Использование вероятностных нейронных сетей при сопоставлении тематической структуры текстов

В основе классификации образцов в вероятностных нейронных сетях (сеть PNN) положено использование метода Байеса. Правилом определения принадлежности образца к одному из заранее сформированных классов является выражение:

$$O \in K \forall (h_k \times c_k \times F_K(x)) = max, K \in \{N\}, \quad (7)$$

где O – классифицируемый образец; K – класс к которому принадлежит образец; $\{N\}$ – множество классов; h_K – априорная вероятность принадлежности образца к классу K; c_K – цена ошибки классификации для класса K; $F_k(x)$ – функция плотности распределения вероятности для класса K; x – область на которой определены классифицируемые образцы.

В задаче распознавания спама априорную вероятность, а также цену ошибки классификации можно выбрать одинаковыми для всех классов (спам, нейтральные письма и целевые письма). Функцию плотности распределения вероятности рекомендуется оценивать с помощью метода Парцена, в котором в качестве ядра используется функция Гаусса [7].

Сеть состоит из входного слоя, слоя образцов, слоя суммирования и выходного слоя. Параметры сети PNN определяется следующим образом:

- Число входных элементов равно числу параметров, которые характеризуют образец.
- Число элементов слоя образцов равно числу учебных образцов.
- Число элементов слоя суммирования равно числу классов.
- Выходной слой состоит из одного выходного элемента (ВЭ).

Весовые значения связей входного слоя и слоя образцов устанавливаются равными элементам соответствующего вектора-образца.

Активность любого элемента слоя образцов, при подаче сигнала от неизвестного образца, определяется в соответствии с выражением:

$$\alpha_j = exp\left(\frac{\sum_{i=1}^{N}(w_{i,j} - x_i)^2}{\sigma^2}\right), \quad (8)$$

где α_j – активность j-го элемента слоя образцов; N – количество входных элементов; x_i – значения i-го параметра классифицируемого образца; w_{ij} – вес связи от i-го элемента входного слоя к j-му элементу слоя образцов; σ – определяемый эмпирически параметр, который задает ширину функции Гаусса (рекомендуемое значение 0,1).

К любому элементу слоя суммирования идут связи только от элемента слоя образцов, принадлежащих соответствующему классу. Весовые значения связей, идущих от элементов слоя образцов к элементам слоя суммирования равны 1. Элементы слоя суммирования

складывают выходные значения элементов слоя образцов. Эта сумма является оценкой значения функции плотности распределения вероятностей для совокупности экземпляров соответствующего класса. Выходной элемент указывает элемент слоя суммирования с максимальным значением активности, т.е. указывает класс, к которому принадлежит образец. Таким образом, все параметры сети PNN определяются непосредственно учебными данными. Поэтому в отличии от перспетронов сети PNN не нуждаются в обучении.

После того, как сеть построена, параметры классифицируемого образца подаются на вход сети. В результате прямого прохода сигнала через сеть выходной слой укажет класс, к которому вероятнее всего принадлежит образец.

Возможная архитектура сети PNN для определения принадлежности образца к одному из двух классов A или B показана на рис.2. В рассматриваемом примере образец характеризуется тремя параметрами, поэтому входной слой состоит из трех элементов. Слой образцов состоит из трех элементов. В данном случае предполагается, что в учебном наборе два образца, принадлежащие классу A и один образец, принадлежащий классу B.

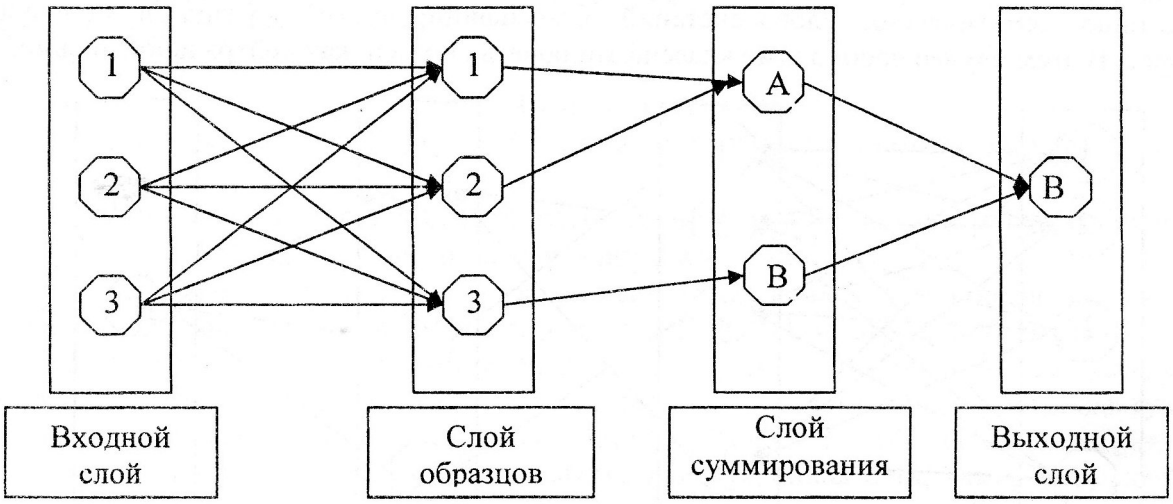

Рис. 2. Пример сети PNN

Рассмотрим применение сети PNN при решении задачи классификации тематической структуры текстов. В первом приближении можно считать, что тематическая структура текста определяется с помощью ограниченного количества словосочетаний, характеризующих содержание тем [5]. Исходя и этого, входными параметрами сети будут указанные тематические словосочетания, приведенные к некоторой стандартной грамматической форме. Обучающая выборка должно состоять с тематических словосочетаний, соответствующих целевым письмам, спаму и нейтральным письмам. Хотя на практике некоторые тематические словосочетания будут присутствовать в разных письмах, но в принципе каждый образец (письмо) может характеризоваться своим уникальным набором тематических словосочетаний. Каждому уникальному словосочетанию и всем его семантическим синонимам, взятым из грамматического словаря, поставим в соответствие отдельный элемент входного слоя. Каждому экземпляру письма из обучающей выборки будет соответствовать отдельный элемент в слое образцов. Таким образом, количество элементов входного слоя (N) будет равно количеству уникальных тематических словосочетаний и их семантических синонимов из обучающей выборки. Количество элементов слоя суммирования (L) равно количеству образцов писем из обучающей выборки. Отметим, что требуемый объем памяти для хранения всего словаря семантических синонимов русского языке не превышает 20 Гб. Поэтому такая структура сети не является слишком ресурсоемкой и вполне реализуема на вычислительной технике среднего класса.

Если тематическое словосочетание присутствует в образце, то вес связи между соответствующими элементами входного слоя и слоя образцов установим 1, в противном случае, вес связи 0. Слой суммирования будет состоять из трех элементов, соответствующих классам целевых писем (*C*), спама (*F*) и нейтральных писем (*S*). Пример рассмотренной структуры сети показан на рис. 3.

Важное отличие нашей сети от классической сети PNN заключается в функционировании элемента выходного слоя. Кроме распознавания элемента слоя суммирования с максимальным значением активности, выходной элемент должен распознавать ситуацию, когда элементы слоя суммирования имеют одинаковый положительный или нулевой уровень активности.

Одинаковый положительный уровень активности может возникнуть, если тематические словосочетания с одинаковой вероятностью встречаются в разных классах, например в спаме и в целевых письмах. Такое письмо возможно классифицировать исходя из презумпции невиновности, т.е. как нейтральное письмо.

Нулевой уровень активности всех элементов слоя суммирования характерен отсутствию тематических словосочетаний классифицируемого образца в обучающей выборке. В этом случае необходимо классифицировать образец, как нейтральное письмо.

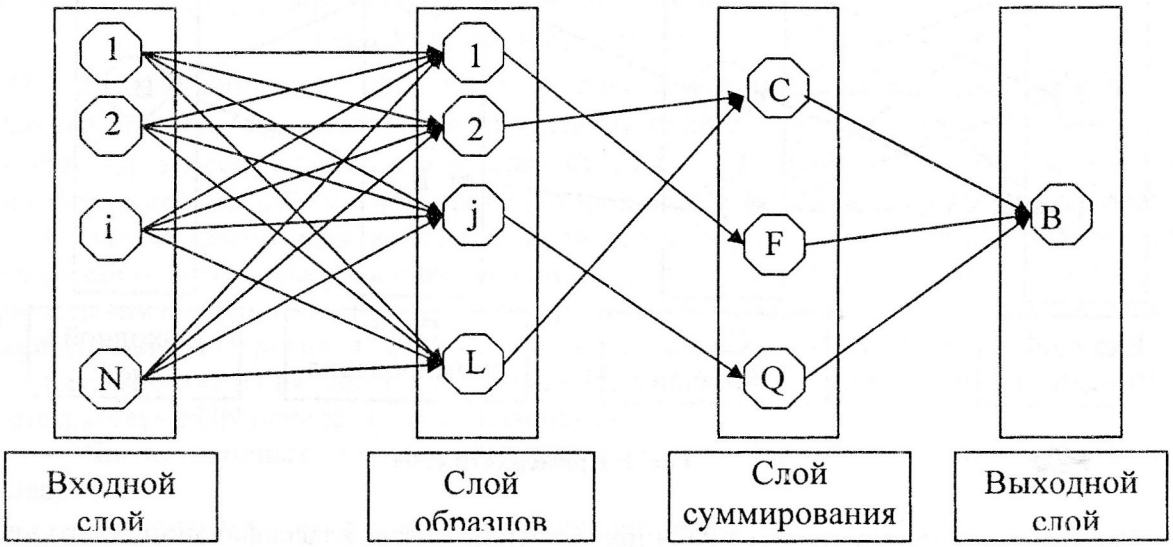

Рис. 3. Пример структуры сети PNN, адаптированной к решению задачи классификации тематической структуры текстов

Важной особенностью реализации рассмотренной сети должна быть возможность добавления новых элементов в входной слой и в слой образцов. За счет этого сеть получит возможность дообучения в процессе эксплуатации на новых образцах. Отметим, что решение этой проблемы не составляет больших трудностей за счет применения объектно-ориентированной технологии создания программного кода.

Выводы

Разработана методика распознавания спама, основой которой является смысловое сопоставление классифицируемого электронного письма с тематикой спама и интересами получателя электронной почты. В методике предусмотрено:

– Применение методов реферирования при извлечении смысловой нагрузки из электронных писем, а также текстов характеризующих интересы.

– Сопоставление результатов реферирования с помощью вероятностной нейронной сети.

Сформирована структура и алгоритм функционирования вероятностной нейронной сети для сопоставления результатов реферирования.

Основными достоинствами систем защиты от спама реализованных на базе данной методики являются:

– Потенциально высокая достоверность классификации электронных писем, в том числе и предназначенных для компрометации спам-фильтров.

– Возможность обучения, как в процессе предварительной настройки, так и в процессе эксплуатации.

– Возможность использования при обучении в процессе предварительной настройки пользователем статистических данных (электронных писем) собранных другими пользователями.

Перспективные пути исследований

– Адаптация существующих методов реферирования к применению в системах распознавания спама.

– Разработка методики сопоставления рефератов на базе рекуррентной семантической нейронной сети.

Список литературы

1. *Цветков В.Я. Булгаков С.В.* Спам и некоторые методы борьбы с ним. – http://vio.fio.ru.

2. *Карташов И.* Революционный метод борьбы со спамом. – http://www.computerra.ru

3. *Спам 2004*: аналитический отчет. – http://www.ashmanov.com.

4. *Заболеева-Зотова А.В.* Естественный язык в автоматизированных системах. Семантический анализ текстов: Монография / ВолгГТУ . – Волгоград, 2002. – 228 с.

5. *Ермаков А.Е.* Эксплицирование элементов смысла текста средствами синтаксического анализа-синтеза. Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2003. – М.: Наука, 2003.

6. *Дударь З.В., Шуклин Д.Е.* Семантическая нейронная сеть, как формальный язык описания и обработки смысла текстов на естественном языке. – Радиоэлектроника и информатика. – Х.: Изд-во ХТУРЭ, 2000. - №3. С. 72-76.

7. *Каллан Р.* Основные концепции нейронных сетей. : Пер. с англ. – М.: Вильямс, - 2003. – 288 с.

Поступила 30.08.2006