

Из достоинств выбора именно технологии Bluetooth является исключение сотовых операторов в участии соединения с серверами для доступа к информационным обучающим ресурсам. Достоинства - автоматическое составление отчётов и списков посещаемости, достаточная безопасность и стабильность соединения. Не надо использовать лишние затраты на оборудование для обучаемых - можно использовать персональные мобильные телефоны, коммуникаторы, карманные компьютеры, ноутбуки и нетбуки.

В будущем возможна организация беспроводной системы на территории факультета радиофизики и электроники на основе технологии беспроводных технологиях передачи данных Bluetooth и Wi-Fi. Данная система будет позволять студенту по приезду на территорию факультета получать свежую и актуальную информацию (расписание занятий на сегодня, важные события на сегодня, текущая успеваемость), сигнализировать о начале учебных занятий, предоставлять доступ в Интернет и к электронным материалам факультета (конспектам лекций, электронного варианта учебных материалов, методических указаний и пр.) Локализованность территории факультета является достоинством для организации такой системы.

#### Литература

1. *Banks, D. A.* Audience response systems in higher education : Applications and cases. Hershey, PA : Information Science Publishing. 2006.
2. *Fies, C.* Classroom response systems : A review of the literature // Journal of Science Education and Technology. 2006. Vol. 15(1), P. 101-109.
3. *Brown, J. S.* Learning, working and playing in the digital age, Washington. 2005.
4. *Wenger, E.* Communities of practice : Learning, meaning and identity. Cambridge University Press. 2002.

## АЛГОРИТМЫ КЛАССИФИКАЦИИ СООБЩЕНИЙ ЭЛЕКТРОННОЙ ПОЧТЫ

**А. А. Семиков**

Белорусский государственный университет, Минск, Беларусь

При проверке сообщений электронной почты большинство пользователей в своих почтовых ящиках регулярно обнаруживают значительное число нежелательных сообщений, обычно относящихся к так называемому «спаму». Существуют различные методы борьбы со спамом, среди которых наиболее распространены метод черных списков и метод классификации. В свою очередь известны различные принципы построения классификаторов сообщений. В данной работе

рассматриваются особенности реализации байесовских и нейросетевых классификаторов применительно к сообщениям электронной почты.

**Байесовский классификатор.** Простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости, называют наивным байесовским классификатором [1].

В случае применения наивного байесовского классификатора к задаче классификации электронных писем на два класса: спам (S) и не-спам ( $\neg S$ ) считается, что документы выбраны из нескольких классов документов одинаковой длины, каждый из которых представлен множеством слов. Предполагается, что вероятность появления  $i$ -го слова в документе независима от длины документа и зависит от класса  $C$ . Вероятность для данного документа  $D$  и класса  $C$  определяется как

$$P(D|C) = \prod_{i=1}^n P(z_i|C)$$

где  $P(z_i|C)$  – вероятность появления  $z_i$ -го слова в документе.

Далее определяется вероятность того, что данный документ  $D$  принадлежит классу  $C$ . Показано, что при

документ может быть отнесен к спаму, в противном случае считается, что это не спам.

На практике байесовские фильтры используются как один из компонентов антиспамовых фильтров [2].

**Нейросетевой классификатор.** Данный классификатор в зависимости от метода обучения может быть реализован в виде персептрона или машины опорных векторов [3, 4]. В рамках данной работы на базе метода опорных векторов реализован и протестирован алгоритм, суть которого заключается в следующем:

- формируется обучающая выборка из ранее полученных сообщений с отметками о принадлежности к тому или иному классу;
- обрабатывается содержимое текстового письма;
- подстраиваются веса с целью увеличения вероятности правильной классификации.

Тестирование классификатора на 15 000 сообщений, среди которых содержались как спам-сообщения, так и нужные, показало, что правильно классифицировалось не более 97% сообщений [5].

Принимая во внимание существование методов, дающих более высокий процент успешной фильтрации, метод опорных векторов оказался недостаточно эффективным.

Обученные на большой выборке, байесовские фильтры способны классифицировать до 99% поступающих писем, причем практическая реализация данного метода достаточно проста. Несмотря на это, существует достаточное количество методов так называемого «отравления» сети. Дело в том, что алгоритм основан на том суждении, что чем больше в письме встречается слов, характерных для нежелательных сообщений, тем с большей вероятностью можно отнести то или иное письмо к числу нежелательных. Легко заметить, что, если это суждение неверно, фильтр перестает работать корректно. На этом основаны нейросетевые методы «обмана» фильтров. Достаточно включить в спам-письма наборы слов, наверняка имеющих положительные веса, и, зная доставлено письмо адресату или попало в карантин (что можно довольно легко проверить), обучить специально созданную нейросеть для эффективного сужения круга включаемых в письмо слов.

Использование слишком строгих фильтров спамовых сообщений приводит к резкому возрастанию вероятности ложной тревоги и потере важных электронных писем.

Следует отметить, что применение алгоритмов классификации сообщений не ограничено проблемой фильтрацией спама. Рассмотренные методы эффективны при их использовании для автоматического распределения сообщений по категориям.

#### Литература

1. *Domingos, P.* On the optimality of the simple Bayesian classifier under zero-one loss // *Machine Learning*, 1997. Vol. 29. P. 103-137.
2. *Ghosh, M.* Exploration of Neuro-Fuzzy Spam Filtering based on Naive Bayesian Filters // *Proc. of the Int. Symp. Knowledge Based Systems*. Las Vegas, 2004.
3. *Bishop, C. M.* *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press, 1977.
4. *Хайкин, С.* *Нейронные сети : полный курс*. Вильямс. 2006.
5. *Семиков, А. А.* Классификация и фильтрация нежелательных email-сообщений с помощью нейросетевых алгоритмов // *Системный анализ и информационные технологии*. Мат. XI Междунар. науч.-тех. конф. (26-29 мая 2009 г., Киев). С. 385.